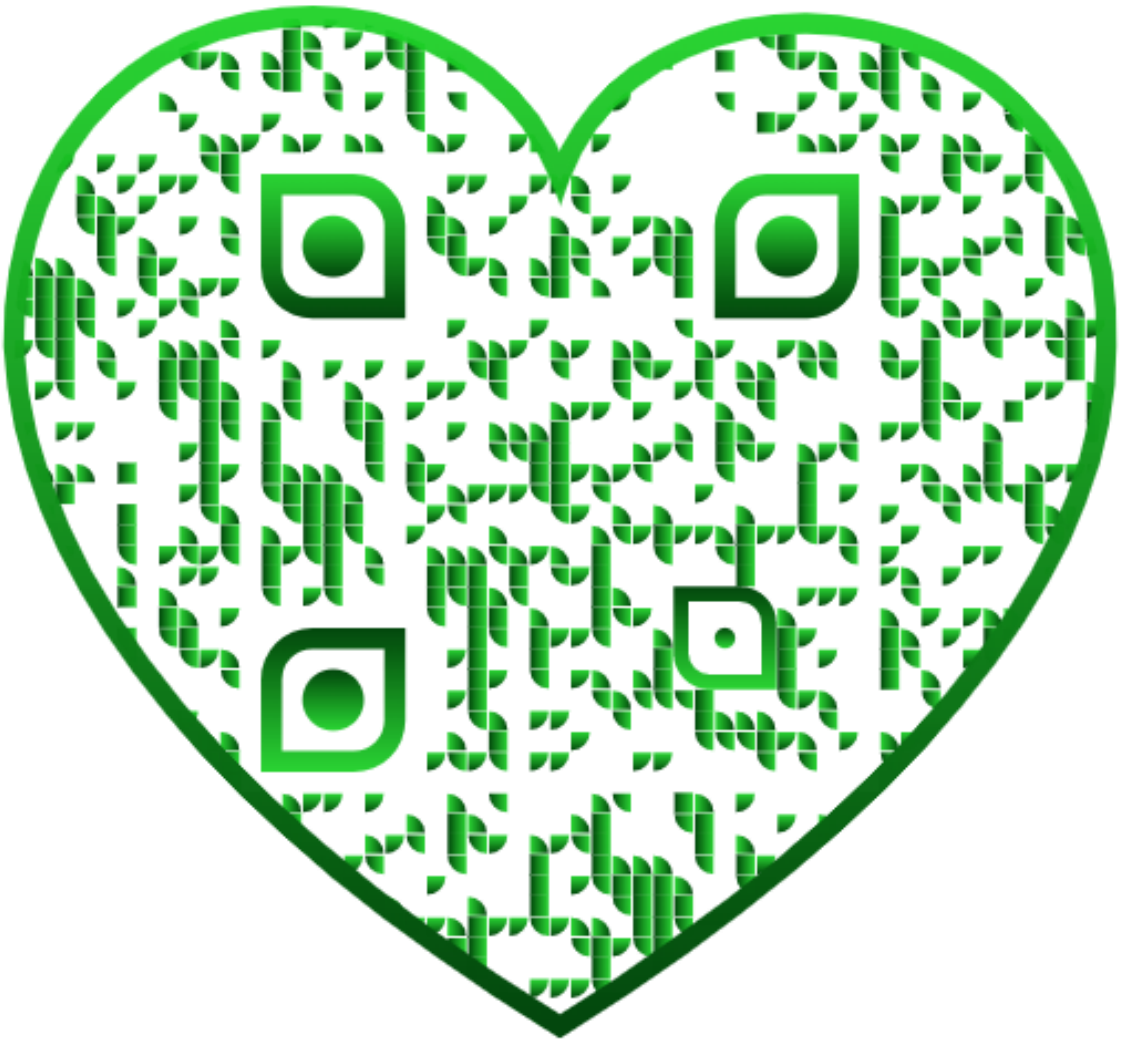
*"Les ordinateurs seront-ils un jour aussi intelligents que les humains ?"*
*- "Oui, mais juste pendant un bref instant". -* **Victor Vinge**

# INTELLIGENCE ARTIFICIELLE

## L'ULTIME RÉVOLUTION

***VERS LA PROSPÉRITÉ OU L'EXTINCTION***

GAËTAN SELLE



www.bookys-gratuit.com

*"Les ordinateurs seront-ils un jour aussi intelligents que les humains ?"*
*- "Oui, mais juste pendant un bref instant". -* **Victor Vinge**

# INTELLIGENCE ARTIFICIELLE

## L'ULTIME RÉVOLUTION

***VERS LA PROSPÉRITÉ OU L'EXTINCTION***

GAËTAN SELLE

Intelligence Artificielle : L’ultime révolution

*Vers la prospérité ou l’extinction*

Gaëtan Selle

Auteur/réalisateur, vidéo généraliste.

Co fondateur de “The Flares”.

The Flares a pour rôle d'éclairer notre chemin vers l’avenir. À travers des documentaires, des formats courts sur YouTube et des articles, nous explorons les théories scientifiques, sociologiques et les avancées technologiques d’aujourd’hui, afin de dresser des hypothèses sur le monde de demain.

www.bookys-gratuit.com

www.the-flares.com

https://www.youtube.com/c/TheFlares

THE FLARES

©2019 Gaëtan Selle - Tous droits réservés

Publié via Amazon Kindle Direct Publishing

Les images utilisées dans ce livre sont soit sous licence “creative common”, soit dans le domaine public.

lsf

2

Du même auteur

*The Mars reality show et autres avenirs (absurdes) possibles.*

*Format Kindle - 2018*

3

Table des matières

Titre

Table des matières

Préface

1. L’Intelligence Artificielle (IA), c’est quoi ?

1.1 Brève historique du domaine

1.2 Qu’est ce que l’intelligence ?

1.3 L’intelligence artificielle dans la culture populaire

1.4 Les types d’intelligence artificielle

Intelligence artificielle limitée (ou faible)

Intelligence artificielle générale (ou forte)

Super intelligence artificielle

2. Aujourd’hui : Intelligence artificielle limitée

2.1 Comment sont faites les IA ?

Programmation

Machine learning (Apprentissage automatique)

Deep learning (Apprentissage profond) et Neural network (réseaux de neurones artificiels)

2.2 Les progrès fulgurants

Résoudre un Rubik’s cube sans aide initiale

Reconnaissance visuelle

Reconnaissance audio du langage

Voitures autonomes

Alpha Go

Jeux vidéo

2.3 Un monde sous les IA limitées

Automatisation et emploi

Créativité

Médecine

Assistants personnels

Armes autonomes

3. Demain : Intelligence artificielle générale (forte)

3.1 Pourquoi est-ce si dur ?

3.2 Les moyens pour y arriver

Une puissance de calcul similaire à celle du cerveau

Simuler le cerveau humain

4

Simuler la sélection naturelle

Comment mesurer l'achèvement d’une intelligence artificielle générale ?

3.3 La question de la conscience

3.4 Les problèmes éthiques à venir

Les 23 principes d'Asilomar sur l’intelligence artificielle

L’éthique des machines

La roboéthique

4. Après demain : Super intelligence artificielle

4.1 L’explosion d’intelligence

4.2 Le problème du contrôle

Le bouton “Stop”

Prison virtuelle

L’alignement des valeurs

Le problème de gouvernance

4.3 Un risque existentiel

Les motivations d’une super intelligence artificielle

Le scénario du roi Midas

4.4 La clé pour un futur viable

Résoudre nos problèmes les plus difficiles

Un oracle

Un génie

Un souverain

4.5 Post humanisme et futur lointain

Post-humanisme

Le futur cosmique de l’humanité

Intelligence sans conscience

5. Quel futur souhaitons nous ?

Les scénarios possibles

Le dictateur bienveillant

Un Dieu protecteur

Fusion post-humaine

Ultime totalitarisme

Perte de contrôle et extinction

Mot de la fin

Sources

5

*À ma famille, qui croit toujours en moi sans forcement comprendre tout ce que je fais ...*

6

# Préface

L'intelligence artificielle est un sujet qui me passionne depuis plusieurs années. Le premier souvenir mémorable associé à ce sujet remonte à 1999. J'avais dix ans, et je suis allé voir un film au cinéma du coin. Après avoir assisté à deux heures de Kung Fu, de science-fiction post apocalyptique, de fusillade, bullet time et lutte contre des machines hyper intelligentes, j'en suis ressorti complètement retourné. Vous avez peut-être deviné qu'il s'agit du film "The Matrix". Ce classique du sous-genre cyberpunk m'a clairement transmis un goût prononcé pour la science-fiction. Et a également planté deux questions dans mon esprit :

- Arrivera-t-on un jour, à concevoir des simulations informatiques indissociables de la réalité ?

- Est-ce que les machines pourraient devenir suffisamment complexes pour surpasser l'intelligence humaine ?

Quelques années plus tard, à mesure que la technologie s'est répandue dans mon quotidien, j'ai pris conscience que ces deux questions allaient très probablement être résolues durant mon vivant. C'est en soi remarquable pour retenir mon attention, mais plus important, cela a d'énormes implications pour le futur de notre espèce, de la vie sur Terre et ça en dit long sur l'accélération du progrès que nous sommes en train de vivre depuis le début du 21e siècle.

Bon, l'idée n'est pas non plus de vous raconter ma vie, mais plutôt d'expliquer le pourquoi du comment derrière ce livre. Je fais donc un petit saut dans le temps jusqu'à 2015. C'est en effet l'année où Marc Durand et moi avons décidé de nous lancer dans la création d'un blog parlant du futur de l'humanité. Face à l'accélération technologique, il nous paraissait important de renseigner le plus de personnes possible sur les différents futurs possibles qui se profilent à l'horizon. Notre intérêt s'est tourné de plus en plus vers une chaîne YouTube où l'on produit des mini documentaires et c'est peut-être là que vous nous avez rencontrés pour la première fois.

Lorsque j'ai décidé d'écrire un ebook sur le futur, le sujet fut évident. Cela ne pouvait être lié qu'à l'une des deux questions semées dans mon esprit par le film "The Matrix". J'ai donc choisi l'intelligence artificielle, car c'est selon moi la conversation 7

la plus importante que l'on puisse avoir collectivement dans cette première moitié du 21e siècle.

C'est à la fois la plus grande promesse, mais également le plus grand danger dans l'histoire de l'humanité. C'est même la chose la plus importante dans l'histoire de la Vie, c'est donc crucial d'en parler. Et j'espère que vous comprendrez pourquoi à mesure que vous lirez ce livre.

Maintenant, je préfère être tout à fait honnête envers vous. Je ne suis pas un chercheur en intelligence artificielle. Je n'ai jamais écrit de thèse. Je n'ai jamais programmé une IA, ni écrit de publication scientifique. Je n'ai jamais étudié les mathématiques. Je ne suis pas non plus un philosophe, un historien, un scientifique, un enseignant ou un futurologue de profession.

Je suis juste un être humain curieux qui cherche à comprendre où l'on va en absorbant beaucoup de contenu à droite à gauche. (Quand je dis beaucoup, c'est beaucoup !).

Si vous cherchez de la crédibilité et de l'autorité sur le sujet, alors ce n'est surement pas le livre pour vous. Vous pensez peut-être que je ne suis qu'un vulgarisateur grossier qui ne sait pas de quoi il parle … et je ne vous en veux pas.

90% du contenu de ce livre vient de conférences que j'ai regardées, de livres et articles que j'ai lus, de podcasts que j'ai écoutés, d'anciens articles que j'ai écrits et bien sûr … Wikipedia. J'avoue même avoir fait beaucoup de copier-coller, c'est dire !

Donc au final, je vois plutôt ce livre comme une tentative de synthétiser de nombreuses idées et recherches plutôt qu'un livre proposant de nouvelles théories résultant de travaux personnels.

Les personnes qui m'ont le plus influencé sont les philosophes Nick Bostrom, Sam Harris, l'historien Yuval Noah Harari, les scientifiques/chercheurs Stuart Russell, Max Tegmark, Lex Fridman, Eliezer Yudkowsky, les entrepreneurs Ray Kurzweil, Elon Musk, Peter Diamandis, les écrivains Isaac Asimov, Philip K.Dick, Tim Urban et les YouTubeurs Isaac Arthur, Two Minutes Paper et Joe Scott.

Les progrès dans ce domaine sont fulgurants, ce qui a rendu la rédaction bien plus compliquée que je ne l'imaginais. Tout simplement, car au moment de boucler un chapitre, il fallait que j'y inclue les nouvelles percées sorties entre temps. J'ai commencé l'écriture le 23 juillet 2018 ce qui m'a permis de me rendre compte de la rapidité des avancées en intelligence artificielle en l'espace de 6-7 mois. Si bien que si vous lisez ce livre en 2020,

de nombreuses parties ne sont peut-être plus à jour. Mais 8

une bonne moitié du contenu décrit plutôt des problématiques sur un futur à moyen terme, ce qui devrait allonger la durée de vie de cet ouvrage … sauf si les machines prennent le contrôle plus vite que prévu !

Avant d'entrer dans le vif du sujet, j'aimerais vous poser quelques questions dont vous êtes libres de répondre ou pas.

1.

Est-ce que vous pensez que l'intelligence artificielle pourrait atteindre le niveau d'intelligence d'un humain dans tous les domaines ?

2.

Est-ce que vous pensez que l'intelligence artificielle pourrait atteindre un niveau d'intelligence qui dépasse complètement celle de l'ensemble des humains dans tous les domaines ?

3.

Est-ce que selon vous, une telle intelligence sera bénéfique ou

catastrophique pour l'humanité ?

Gaëtan Selle

9

1.

L'Intelligence Artificielle (IA), c'est quoi ?

10

1. L'intelligence artificielle (IA), c'est quoi ?

Brève historique du domaine

Le terme “Intelligence artificielle” serait apparu officiellement en 1956 lors d’une conférence dans le New Hampshire aux États unis. Mais le concept d’une machine ou objet capable de raisonnement et de pensée complexe remonte bien plus loin dans l’Histoire de l’humanité. C’est notamment à travers les anciens mythes que l’on trouve le plus de récits illustrant cette notion. Comme avec le mythe de Talos : un automate géant fait de bronze ayant pour but de protéger Europe, mère du Roi de Crète. La statue de Pygmalion qui prend vie ou encore dans la mythologie juive, avec le Golem, humanoïde fait d’argile qui se voit insuffler la vie afin de défendre son créateur. C’est à croire que l’idée de créer la vie et élaborer une copie de nous même est profondément ancrée dans la psyché humaine.

Le géant Talos armé d’une pierre sur une pièce de Phaistos, Crète

- 280-270 av. J-C.

Mais les philosophes de nombreuses civilisations se sont intéressés à l’idée de comprendre la pensée humaine par une approche mécanisée, ce qui implique la possibilité de la reproduire. C’est le début de la logique et du raisonnement. Au 4e siècle av. J.-C., Aristote (384 av. J.-C - 322 av. J.-C.) invente un système de raisonnement logique appelé “Syllogisme”. C’est important, car la programmation informatique est un prolongement de l’approche mathématique de la

logique. Le syllogisme fonctionne en mettant en relation trois propositions afin d'en déduire un résultat. Un exemple très connu : "*Tous les hommes sont mortels, or Socrate est un homme donc Socrate est mortel*".

Le mathématicien et astronome Arable Al-Khwarizmi's (780 – 850) fonda la discipline de l'algèbre et donna son nom au terme "Algorithme" au 9e siècle apr. J.-C..

12

Le philosophe espagnol Roman Llull (1232–1315) est considéré comme un pionnier du calcul informatique par le système de classification qu'il a inventé. L'innovation radicale que Llull a introduite dans le domaine de la logique est la construction et l'utilisation d'une machine faite de papier pour combiner des éléments de langage. À

l'aide de figures géométriques connectées, suivant un cadre de règles défini avec précision, Llull a essayé de produire toutes les déclarations possibles dont l'esprit humain pouvait penser.

Au 17e siècle, plusieurs philosophes explorent la possibilité que la pensée puisse être expliqué par un ensemble d'opérations algorithmiques ce qui fait de l'être humain et les animaux, ni plus ni moins que des machines complexes. C'est le cas avec Leibniz (1646 - 1716) qui a spéculé que la raison humaine pourrait être réduite à un calcul mécanique et à qui on doit la première utilisation d'un système binaire. Mais aussi Renée Descartes (1596 - 1650) et Hobbes (1588 - 1679) avec sa célèbre citation

: "*reason is nothing but reckoning*" que je traduis par : "*La raison n'est rien d'autre que du calcul mathématique*".

Au 18e siècle, les jouets mécaniques sont très populaires. En 1769, l'ingénieur hongrois Baron Wolfgang von Kempelen (1734 - 1804) construisit une machine à jouer aux échecs pour amuser la reine autrichienne Maria Theresia. C'était un appareil mécanique, appelé le Turc (car ressemblant à un homme turc traditionnel de l'époque).

Un des automates les plus célèbres de l'histoire, mais qui étaient en réalité un canular bien ficelé. Toutefois, cela montre la fascination de construire des machines capables de s'adonner à des activités intellectuelles comme les échecs.

13

Reconstruction du Turc - Image de Carafe at English Wikipedia.

CC BY-SA 3.0

En 1818, Marry Sheley (1797 - 1851) publie l’une des premières histoires de science-fiction, “Le monstre de Frankenstein”, qui raconte l’histoire d’une créature conçue de toute pièce. Cette dernière prend vie et devient folle, se rebellant contre son créateur. Une thématique reprise de nombreuses fois par la suite, surtout dans le domaine de l’intelligence artificielle, robots et machines.

Au 20e siècle, l'étude de la logique mathématique a fourni la percée essentielle qui a rendu l'intelligence artificielle plausible. Les acteurs majeurs étant Bertrand Russell et Alfred North Whitehead avec l’ouvrage “Principia Mathematica” en 1913. La première machine fonctionnelle capable de jouer aux échecs remonte à 1912 avec 14

l’invention de “Ajedrecista” par Leonardo Torres. Le terme “robot” est utilisé pour la première fois dans une pièce de théâtre à Londres en 1923, écrit par Karel Capek (1890 - 1938).

Alors bien sûr, l’histoire de l’intelligence artificielle côtoie étroitement celle de l’informatique. Le 20e siècle peut donc surement être considéré comme le siècle qui a vu naître cette discipline. Toutefois, il ne faut pas oublier que des machines à calculer ont été construites depuis l'Antiquité et améliorées à travers l'histoire par de nombreux mathématiciens. Le mécanisme d'Anticythère est le premier "ordinateur" analogique et mécanique, datant d’au moins 2000 ans. Charles Babbage (1791 - 1871) et Ada Lovelace (1815 - 1852) sont souvent considérés comme les premiers pionniers de l'informatique au début du 19e siècle. Ils ont développé les plans pour un ordinateur analogue appelé la machine analytique. Ada Lovelace formula pour la première fois la notion de programme au cours de son travail avec le mathématicien Charles Babbage.

Mais la machine analytique ne fut fabriquée qu'après leur mort.

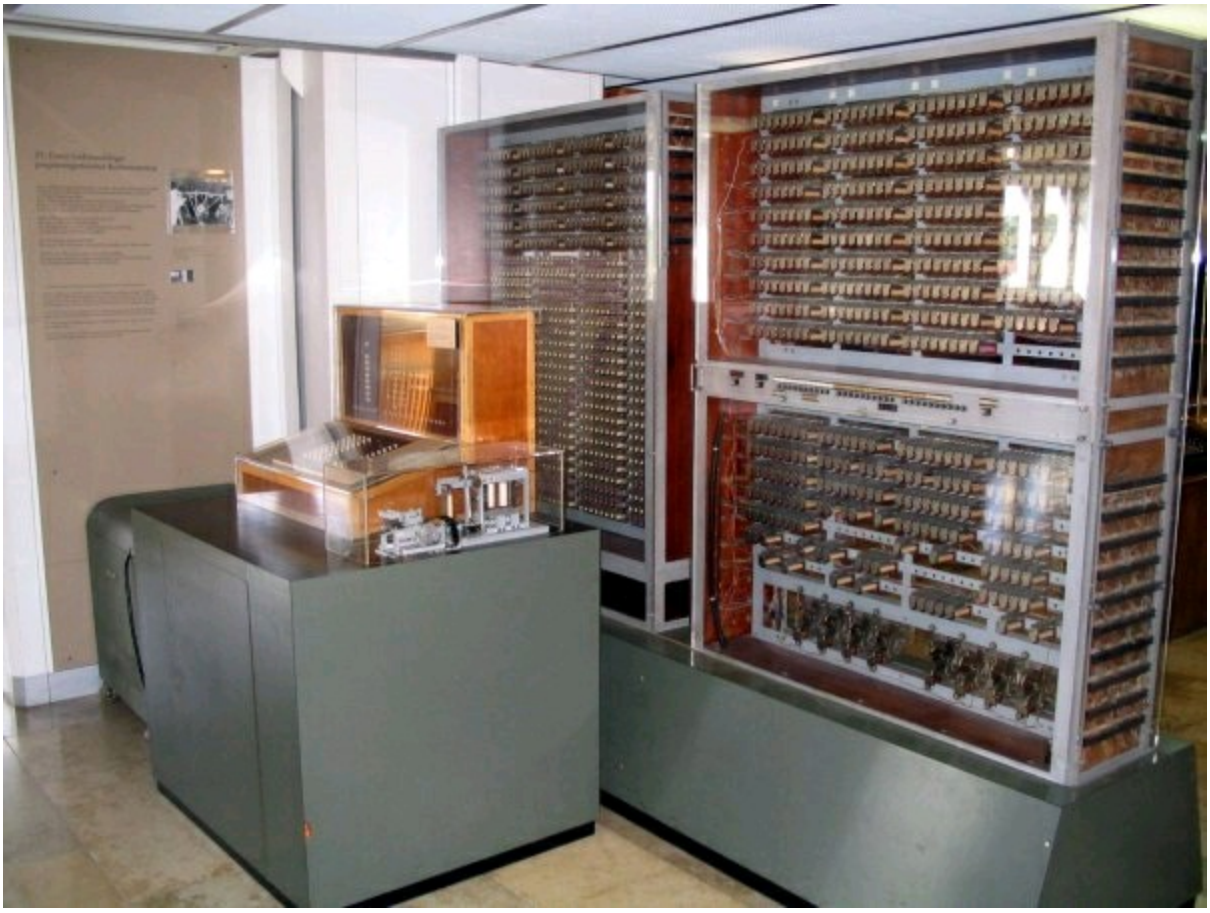
Portrait d’Ada Lovelace - 1840.

On fait un saut dans le temps pour assister à la naissance de l’informatique moderne avec un nom qui ressort en particulier : Alan Turing (1912 - 1954). En 1936, Alan Turing et Alonzo Church publient la thèse de Church-Turing, une hypothèse sur la nature des dispositifs de calcul mécanique. La thèse soutient que tout calcul possible peut être effectué par un algorithme exécuté sur un ordinateur, à condition que suffisamment de temps et d'espace de stockage soient disponibles. Les premiers ordinateurs modernes furent les machines destinées à déchiffrer

des codes nazis durant la Seconde Guerre mondiale (ENIAC et Colossus). Un des premiers ordinateurs programmables est attribué à l'ingénieur allemand Konrad Zuse (1910 -

1995) en 1941 avec le Z3.





Réplique du Zuse Z3 au musée Deutsches Museum de Munich.

Le concept de machine pensante se répand dans les années 40 et 50 avec des personnalités comme John Von Neumann (1903 - 1957), Norbert Wiener (1894 -

1964) et Claude Shannon (1916 - 2001). En 1950, Alan Turing stipule dans une publication que si une machine peut tenir une conversation (sur un téléscripteur) qui est indiscernable d'une conversation avec un être

humain, il est raisonnable de dire que la machine “pense”. C’est ainsi que naît le célèbre test de Turing. On note aussi en 1950 la première apparition des célèbres trois lois de la robotique par l’auteur de science fiction, Isaac Asimov (1920 - 1992).

Mais c’est en 1956, lors de la conférence de Dartmouth, aux États unis, que l’on peut mettre une date précise sur la naissance du domaine de recherche intitulé :

“Intelligence artificielle”. L'idée maîtresse de la conférence c’est que chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut être décrit avec une telle précision qu'une machine peut être conçue pour la simuler. Et beaucoup de chercheurs appartenant à la première vague de recherche en IA étaient plutôt optimistes. Certains pensaient à la fin des années 50 que d’ici 10 ans, le champion du monde des échecs serait un ordinateur. Ou encore cette affirmation par Marvin Minsky (1927 - 2016) en 1970 : *“D’ici trois à huit ans, nous aurons une machine avec* 17

*l'intelligence générale d'un être humain moyen.* ”

Cet optimisme ne dura pas puisque le domaine de l’intelligence artificielle connut un premier ralentissement de 1974 à 1985. Beaucoup de chercheurs avaient sous-estimé les difficultés et les limites technologiques de l'époque. Mais des notions intéressantes furent étudiées comme l’apprentissage automatique (machine learning) ou les réseaux de neurones artificiels avec le Perceptron (Neural network et deep learning). Deux techniques utilisés aujourd’hui, mais on y reviendra.

L’histoire moderne de l’intelligence artificielle est une sorte de va-et-vient entre enthousiasme et déception. Entre progrès, et glaciation. Étant majoritairement financés par les gouvernements à cette époque, notamment aux États unis par le DARPA (ministère de la Défense), les chercheurs étaient donc dépendant des subventions et si les progrès n'étaient pas assez rapides aux goûts des politiciens, il était très difficile de trouver les financements pour un projet impliquant l’intelligence artificielle. Ce qui a eu pour résultat d’engendrer trois grandes vagues de recherche.

La première commença en 1956, la seconde dans les années 80, et nous sommes actuellement dans la troisième depuis la fin du 20e.

Dès le début des années 80, un nouveau type de programme appelé "Systèmes experts" permit un regain d'intérêt pour l'intelligence artificielle. Les systèmes experts sont conçus pour résoudre des problèmes complexes en raisonnant à travers des amas d'information, représentés principalement comme des règles "si-alors". Les entreprises du monde entier ont commencé à développer et utiliser des systèmes experts et, en 1985, elles dépensaient plus d'un milliard de dollars dans l'intelligence artificielle. Dans le même temps, les différents projets informatiques japonais ont inspiré les gouvernements américain et britannique à rétablir le financement dans la recherche universitaire. Ce fut la deuxième vague de recherche en intelligence artificielle. Pourtant, dès le début des années 90, une nouvelle glaciation toucha le champ d'études de l'intelligence artificielle qui acquit une mauvaise réputation. Jugé irréaliste et donc une perte de temps.

Il faut attendre la fin des années 90 pour finalement voir l'intelligence artificielle sur le devant de la scène avec un événement majeur : La victoire d'un ordinateur face au champion du monde d'échecs, Garry Kasparov le 11 mai 1997. Deep blue était un superordinateur produit par IBM et capable de calculer jusqu'à 200 000 000 de coups par seconde. L'intelligence artificielle a commencé à être utilisé avec succès pour la logistique, l'exploration de données, le diagnostic médical, la robotique industrielle, la reconnaissance vocale, le moteur de recherche de Google et d'autres domaines. En grande partie grâce à une puissance de calcul croissante suivant la loi de Moore, qui est l'observation que le nombre de transistors dans un circuit intégré double environ 18

tous les deux ans pour un coût constant de $1000. Le prix baisse pour une plus grande puissance de calcul. Cette troisième vague de recherche est toujours d'actualité et ne semble pas être menacée par une nouvelle glaciation. Nous sommes désormais à un stade où les technologies ont prouvé concrètement qu'elles sont efficaces et utiles, ce qui encourage les investissements dans ce domaine.

Le nom "Agent intelligent" se répandit pour parler d'un système qui

perçoit son environnement et prend des mesures afin de maximiser ses chances de succès. Selon cette définition, les programmes simples qui résolvent des problèmes spécifiques sont des "agents intelligents". Le début du 21e siècle voit de nombreux problèmes résolus grâce aux algorithmes de l'intelligence artificielle. Pourtant, ce champ d'études ne reçoit pas tous les crédits qu'ils méritent. Toutes les innovations dans le domaine de l'intelligence artificielle sont réduites au statut de gadget de l'informatique.

L'idée de pouvoir créer une intelligence artificielle ayant le niveau d'intelligence humain divisait toujours autant la communauté de chercheurs. Il y avait peu d'accord sur les raisons pour lesquelles nous n'avions pas réussi à réaliser le rêve qui avait captivé l'imagination du monde dans les années 1960. De nombreux chercheurs en IA ont délibérément appelé leur recherche par d'autres noms, tels que systèmes cognitifs ou intelligence computationnelle. En partie pour éviter d'être perçu comme des rêveurs, mais aussi, car les nouveaux noms aidaient à obtenir des financements.

Il faut également mentionner que le terme IA était souvent associé à la science-fiction, notamment au cinéma avec plusieurs succès planétaires (2001 l'odyssée de l'espace, Terminator, Matrix). Le public avait donc une image exagérément fausse du domaine avec cette idée que les machines vont se rebeller dès qu'elles seront intelligentes. Du coup, certains chercheurs préféraient tout simplement éviter le terme.

Au fil du début du 21e siècle et l'explosion d'Internet, l'accès à de grandes quantités de données ont permis des avancées majeures dans l'intelligence artificielle.

En 2011, Watson, conçu par IBM, remporte le Jeopardy, un jeu de question/réponse télévisé, face à deux des meilleurs candidats de l'époque. IBM souhaite que Watson soit utilisé pour diagnostiquer des maladies. En mars 2016, le programme AlphaGo, développé par Google DeepMind, remporte 4 des 5 parties du jeu de Go dans un match avec un champion du domaine, Lee Sedol. Cela marque l'achèvement d'une étape importante dans le développement de l'intelligence artificielle. Go étant un jeu extrêmement complexe, bien plus que les échecs puisque chaque tour place le joueur face à des centaines de millions de combinaisons.

En 2016, le marché des produits, matériels et logiciels liés à l'IA a atteint plus de 8

milliards de dollars. L'apprentissage automatique (Machine learning) et 19

l'apprentissage approfondi (deep learning) ont montré leur force dans la fabrication de système capable de rivaliser avec l'être humain dans de nombreux domaines incluant la conduite de véhicule, la reconnaissance d'image, et le langage. Google, Apple, Microsoft, Facebook, Amazon et les plus grandes entreprises de la Silicon Valley ont toutes misé sur l'intelligence artificielle en les mettant au coeur de leurs futures ambitions.

Si l'on prend du recul sur l'histoire de l'humanité, on remarque qu'il y a de nombreux événements qui ont façonné la trajectoire de notre espèce à travers les époques. Que ce soit les guerres, les changements politiques, émergence de religion, révolutions sociales, technologiques, scientifiques et bien d'autres. Mais si vous êtes pressé, et que vous devez résumer l'ensemble de l'histoire d'Homo Sapiens à une race extra-terrestre, vous pouvez le faire en seulement trois grandes révolutions.

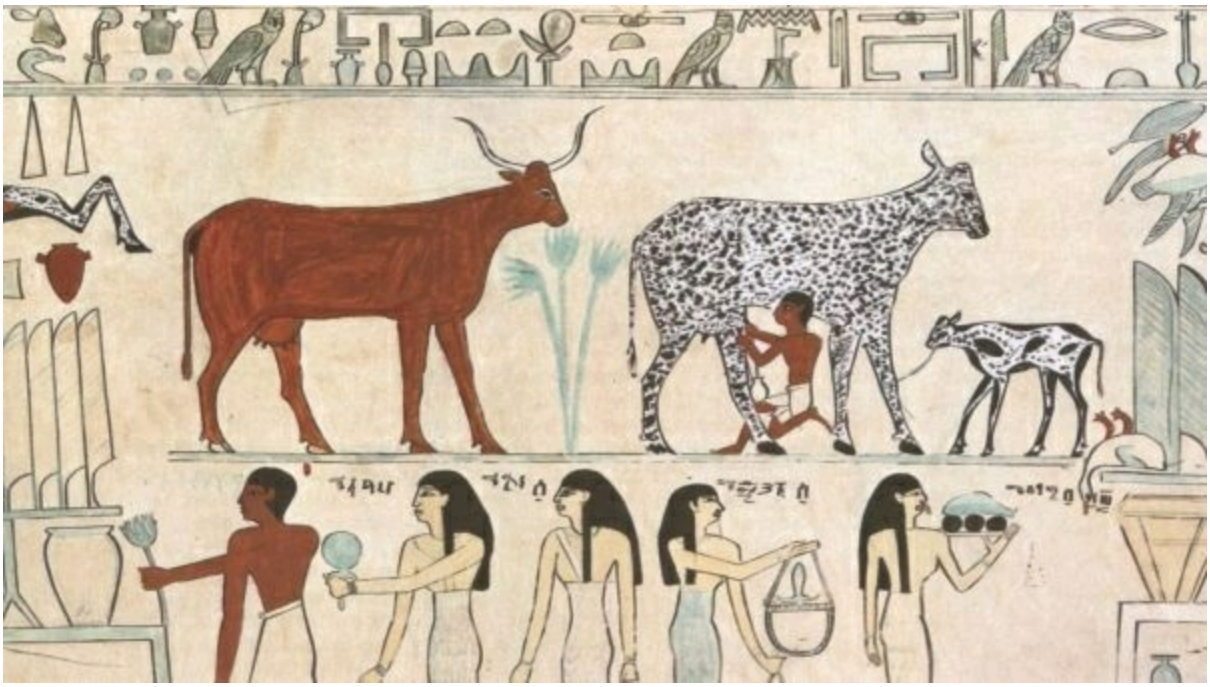
La première c'est la révolution cognitive qui est apparue il y a 100 000 - 70 000

ans. Une espèce insignifiante appartenant à la famille des grands singes et localisée dans un petit coin de l'Afrique de l'Est commença à développer des capacités cognitives uniques comme la pensée abstraite, la mémoire, le raisonnement, la communication en groupe ce qui lui permit de devenir l'animal le plus puissant sur la planète. Homo Sapiens s'est répandu ensuite sur tous les continents et finit par devenir le seul représentant du genre homo.

La seconde grande révolution se déroula il y a environ 12 000 ans. Étant essentiellement des tribus nomades de chasseurs/cueilleurs, des Homo Sapiens du bassin mésopotamien domestiquèrent certains types de plantes et animaux. Ce changement permit la sédentarisation, la construction de cité et l'élaboration de structure sociétale plus

complexe. Cette deuxième révolution, appelée révolution agricole, engendra une explosion de la population.

20

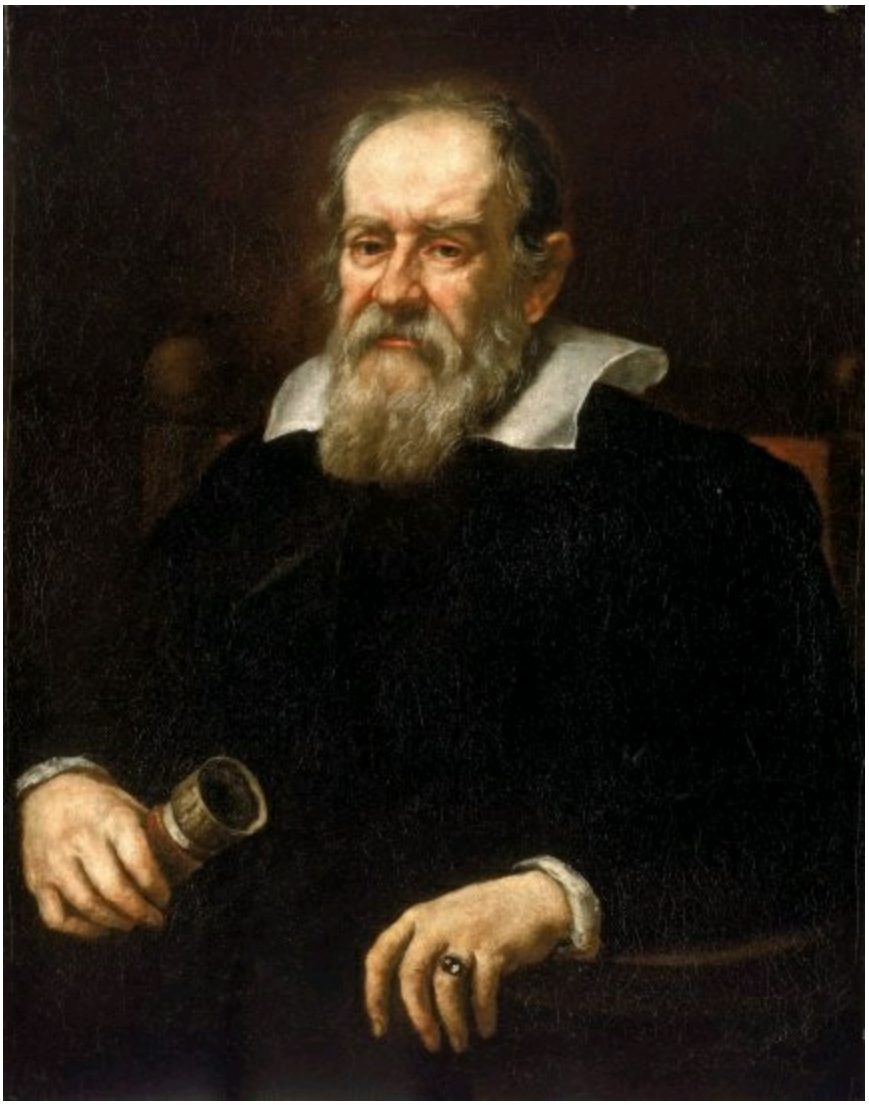


Vaches domestiques dans l’ancienne Egypte.

Et enfin la troisième révolution eut lieu il y a 500 ans principalement en Europe avec la révolution scientifique. Elle a permis à Homo Sapiens de comprendre de plus en plus les lois qui gouvernent le monde autour de lui, mais également à l'intérieur.

Cette compréhension fut la clé permettant la maîtrise de certaines forces de la nature, l’utilisation de nouvelles sources d'énergie et l’exploitation de la puissance des machines. Elle permit la révolution industrielle au 18e siècle qui engendra également une explosion de la population grâce notamment aux progrès de la médecine et à l’agriculture de masse.

21



Portrait de Galileo Galilei par Justus Sustermans -1636.

Mais si on se projette dans un siècle ou deux, notre époque sera peut-être perçue comme le début de la quatrième grande révolution. Une sorte de deuxième révolution cognitive où l'intelligence d'Homo Sapiens sera décuplée par un facteur si grand que cela pourrait bien déboucher sur la naissance d'une nouvelle espèce. La quatrième révolution sera la révolution de l'intelligence artificielle.

**Dates clés :**

Première apparition du terme robot : 1923

Concept du premier ordinateur moderne : 1936

Isaac Asimov invente les 3 lois de la robotique : 1950

Le test de Turing est imaginé : 1950

Première apparition du terme intelligence artificielle : 1956

Un ordinateur bat le champion du monde d'échecs : 1997

Une IA gagne à Jeopardy : 2011

AlphaGo bat un champion du jeu de Go : 2016

22

Première utilisation commerciale d'une flotte de voiture autonome : 2018

lsf

23

1. L’Intelligence Artificielle (IA), c’est quoi ?

Qu’est ce que l’intelligence ?

Pour comprendre ce qu’est une intelligence artificielle, il faut commencer par se demander ce qui constitue l’intelligence. La question peut paraître évidente, mais l’Histoire nous montre que la définition de l’intelligence a évolué au cours des siècles.

Socrate a dit: " *Je sais que je suis intelligent, parce que je sais que je ne sais rien.* "

Einstein quand a lui déclara : " *Le vrai signe de l'intelligence n'est pas la connaissance, mais l'imagination.* " Plus récemment, les neuroscientifiques sont entrés dans le débat, cherchant des réponses sur l'intelligence d'un point de vue scientifique: qu'est-ce qui rend certains cerveaux plus intelligents que d'autres? Les personnes intelligentes sont-

elles plus en mesure de stocker et de récupérer des informations ?

Ou peut-être que leurs neurones ont plus de connexions leur permettant de combiner des idées de manière créative ?

Si on se place purement du point de vue d'un observateur, on remarque qu'en regardant simplement l'étendue de ce que font les êtres humains, il semble que l'on ait plus de compétence générale que les chimpanzés. Les chimpanzés ont plus de compétence que les souris, les souris plus de compétences que les moustiques, et les moustiques plus de compétences que les bactéries. Il y a donc une sorte d'échelle verticale et horizontale de compétence. Toujours en tant qu'observateur, on peut raisonnablement conclure que ce qui fait la différence entre l'étendue des compétences d'une espèce à une autre se trouve dans la complexité de l'organisme. Et plus particulièrement dans ce qui se trouve dans la boîte crânienne. L'intelligence est le facteur qui permet à ces différentes espèces de faire plus de choses que les autres.

D'être plus haut qu'une autre sur l'échelle. L'intelligence peut donc être quantitative, et qualitative.

**Définitions**

Le terme "intelligence" dérive des noms latins intelligentia ou intellēctus, qui à son tour vient du verbe intelligere, signifiant connaître, comprendre ou percevoir. Au Moyen-Âge, intellectus est devenu le terme technique savant pour la compréhension.

Cependant, ce terme était fortement lié aux théories théologiques, y compris les théories de l'immortalité de l'âme. Cette approche a été fortement rejetée par les premiers philosophes modernes tels que Francis Bacon (1561 - 1626), Thomas Hobbes (1588 - 1679), John Locke (1632 - 1704) et David Hume (1711 - 1776). Tous ont préféré le mot "compréhension". Le terme "intelligence" est donc devenu moins 24

commun, mais a été repris plus tard dans la psychologie contemporaine.

En feuilletant les dictionnaires, on remarque que l'intelligence est souvent

définie comme la capacité générale d'apprendre et d'appliquer des connaissances pour manipuler son environnement, ainsi que la capacité de raisonner et faire preuve d'abstraction. Les autres définitions de l'intelligence comprennent la capacité d'évaluer et de juger ou encore la capacité de comprendre des idées complexes. Il semblerait qu'on ait du mal à trouver un consensus.

Si on se tourne sur le dictionnaire Larousse, voici ce que l'on trouve :

*Intelligence* : nom féminin (latin intelligentia, de intelligere, connaître) :

- Ensemble des fonctions mentales ayant pour objet la connaissance conceptuelle et rationnelle : Les mathématiques sont-elles le domaine privilégié de l'intelligence ?

- Aptitude d'un être humain à s'adapter à une situation, à choisir des moyens d'action en fonction des circonstances : Ce travail réclame un minimum d'intelligence.

- Personne considérée dans ses aptitudes intellectuelles, en tant qu'être pensant : C'est une intelligence supérieure.

- Qualité de quelqu'un qui manifeste dans un domaine donné un souci de comprendre, de réfléchir, de connaître et qui adapte facilement son comportement à ces finalités : Avoir l'intelligence des affaires.

- Capacité de saisir une chose par la pensée : Pour l'intelligence de ce qui va suivre, rappelons la démonstration antérieure.

Mais la mesure de l'intelligence est encore plus sujette à controverse et désaccord.

Bien qu'il existe un certain nombre de méthodes pour mesurer l'intelligence, la méthode standard et la plus largement acceptée consistent à mesurer le "quotient intellectuel" (QI) d'une personne. Basé sur une série de tests qui évaluent différents types de capacités telles que mathématique, spatiale, verbale, logique, la mémorisation et qui aboutit à un score.

## Théories psychométriques

L'une des plus anciennes théories sur l'intelligence est venue du psychologue britannique Charles E. Spearman (1863-1945), qui publia son premier article majeur sur l'intelligence en 1904. Il conclut que seulement deux types de facteurs sous-tendent toutes les différences individuelles dans les résultats des tests d'intelligence. Il a appelé le premier "facteur général", ou (g) qui est présent sur toutes les tâches nécessitant une intelligence. Le second facteur est spécifiquement lié à chaque test particulier, noté (s). Par exemple, quand une personne fait un test d'arithmétique, sa performance sur le test nécessite un facteur général qui est commun à tous les tests (g) 25

et un facteur spécifique lié aux opérations mentales qui sont nécessaires pour le raisonnement mathématique (s).

Le psychologue américain LL Thurstone était en désaccord avec la théorie de Spearman, affirmant plutôt qu'il y avait sept facteurs, qu'il identifie comme les

"capacités mentales primaires". Ces sept capacités sont la compréhension verbale, la fluidité verbale, l'arithmétique, la visualisation spatiale, le raisonnement déductif, la mémoire et la rapidité de perception. Le Canadien Philip E. Vernon (1905-1987) et l'Américain Raymond B. Cattell (1905-1998) ont tenté de rapprocher les deux théories en suggérant que les capacités intellectuelles sont hiérarchisées avec l'intelligence (g) au sommet et les capacités mentales primaires, en dessous. Plus tard, Cattell a suggéré que le facteur (g) peut être divisé en intelligence fluide et intelligence cristallisée : alors que l'intelligence fluide est représentée par le raisonnement et la résolution de problèmes, l'intelligence cristallisée est la connaissance acquise au fil des années. Mais vu que les théories psychométriques ne pouvaient pas expliquer les processus produisant l'intelligence, certains chercheurs sont partis dans une autre direction.

## Psychologie cognitive

Elles supposent que l'intelligence comprend un ensemble de

représentations mentales de l'information et un ensemble de processus qui les opèrent. Une personne plus intelligente aura donc une meilleure représentation de l'information et pourra y travailler plus rapidement. Une conception qui fait penser à l'informatique puisque l'on parle d'information et de rapidité de traitement de cette information. Ce qui a donné des idées à certains psychologues cognitifs qui ont étudié l'intelligence humaine en construisant des modèles informatiques de la cognition humaine. Deux leaders dans ce domaine étaient les informaticiens américains Allen Newell (1927 -

1992) et Herbert A. Simon (1916 - 2001). À la fin des années 1950 et au début des années 60, ils ont travaillé avec l'expert en informatique Cliff Shaw (1922 - 1991) pour construire un modèle informatique appelé "General Problem Solver"

(Solutionneur de problèmes généraux). Il pouvait trouver des solutions à un large éventail de problèmes assez complexes, tels que des problèmes logiques et mathématiques. Cependant, la psychologie cognitive n'a pas répondu à la question de savoir pourquoi un certain comportement est considéré comme intelligent.

**Le contextualisme**

Le contextualisme examine comment les processus cognitifs opèrent dans divers contextes environnementaux. Les deux théories les plus influentes de cette approche sont la théorie triarchique de l'intelligence humaine de Sternberg et les intelligences multiples de Gardner. Gardner est allé un peu plus loin que les chercheurs précédents 26

qui ont suggéré que l'intelligence comporte plusieurs capacités. Il a soutenu qu'on ne peut pas parler d'une seule intelligence, mais bien de plusieurs : intelligence linguistique, intelligence logico-mathématique, intelligence spatiale, intelligence intrapersonnelle, intelligence interpersonnelle, intelligence corporelle kinesthésique, intelligence musicale, intelligence naturaliste, intelligence existentielle (ou spirituelle).

**Les théories biologiques**

L’idée étant que la vraie compréhension de l'intelligence n'est possible qu'en identifiant sa base biologique. Les théories biologiques essaient de comprendre les bases neuronales de l'intelligence et non, comme les trois autres approches, les constructions hypothétiques. Cette façon réductionniste de regarder le cerveau, rendu possible par les récents progrès technologiques, semble prometteuse pour construire un modèle partant de la source de l’intelligence, c’est-à-dire le cerveau. Certaines théories suggèrent que l'intelligence pourrait être le résultat de neurones plus efficacement connectés ou de la transmission rapide de l'information à travers les axones des neurones. De même, très récemment, le rôle des cellules gliales, autrefois considérées comme moins importantes, a pris le devant de la scène.

Alors tout cela est très intéressant pour comprendre le pourquoi du comment de notre intelligence, mais ça tourne quand un peu trop autour de notre nombril. Difficile de faire passer un test de QI à des souris. Au final, cela ne nous dit pas si les autres mammifères sont intelligents. Ou encore les oiseaux ? Les insectes ? Et les plantes alors ? Et la question qui nous intéresse le plus par rapport au sujet de cet ouvrage…

Et les machines ?

Il nous faut donc une définition de l’intelligence qui soit plus large afin de ne pas seulement inclure les Homo Sapiens que nous sommes. Si nous devions classer les plantes, les insectes, les poissons, les reptiles, les mammifères ou les machines, sur la base de l'intelligence, quel système de mesure utiliser ?

**Une définition globale de l’intelligence**

Si nous définissons l'intelligence par la durabilité (c'est-à-dire la capacité de survivre sur une longue période de temps), les êtres les plus intelligents pourraient être les arbres ou les plantes en général. Les arbres créent leur propre nourriture, se reproduisent assez efficacement en dispersant leurs graines autour d’eux et certains peuvent vivre dans des environnements extrêmes. Les plantes ont survécu sur Terre plus longtemps que n'importe quel insecte, amphibien, reptile, mammifère ou oiseau.

Les plantes sont-elles les êtres les plus intelligents sur Terre ?

Si nous définissons l'intelligence par la complexité du système de communication, 27

alors les baleines et les dauphins pourraient disposer d'une structure de communication tout aussi complexe et voir même plus complexes que les humains selon certains scientifiques. Est-ce que les dauphins sont les êtres les plus intelligents sur Terre, tout comme dans le roman "le guide du voyageur galactique" ?

Si nous définissons l'intelligence par l'organisation sociale, alors les fourmis et les abeilles sont manifestement intelligentes. Une fourmi, en soi, ça n'a pas l'air très brillant. Mais à l'échelle d'une colonie, les fourmis présentent de nombreuses caractéristiques et comportements que nous associons avec l'intelligence. En fait, si les fourmis n'existaient pas sur Terre, mais, par exemple, sur Mars, je suis sûr que l'on se demanderait si nous n'avons pas rencontré une race extra-terrestre intelligente qui construit des villes, développe l'agriculture, élève des autres espèces, possède des rangs sociaux tels que les princesses, les soldats, les travailleurs et les esclaves. On se dirait forcément que ces extraterrestres sont intelligents.

La définition de l'intelligence telle qu'elle est présente dans la plupart des dictionnaires s'oriente un peu trop vers une description de notre propre intelligente, ce qui nous rend aveugles à l'intelligence que l'on trouve autour de nous.

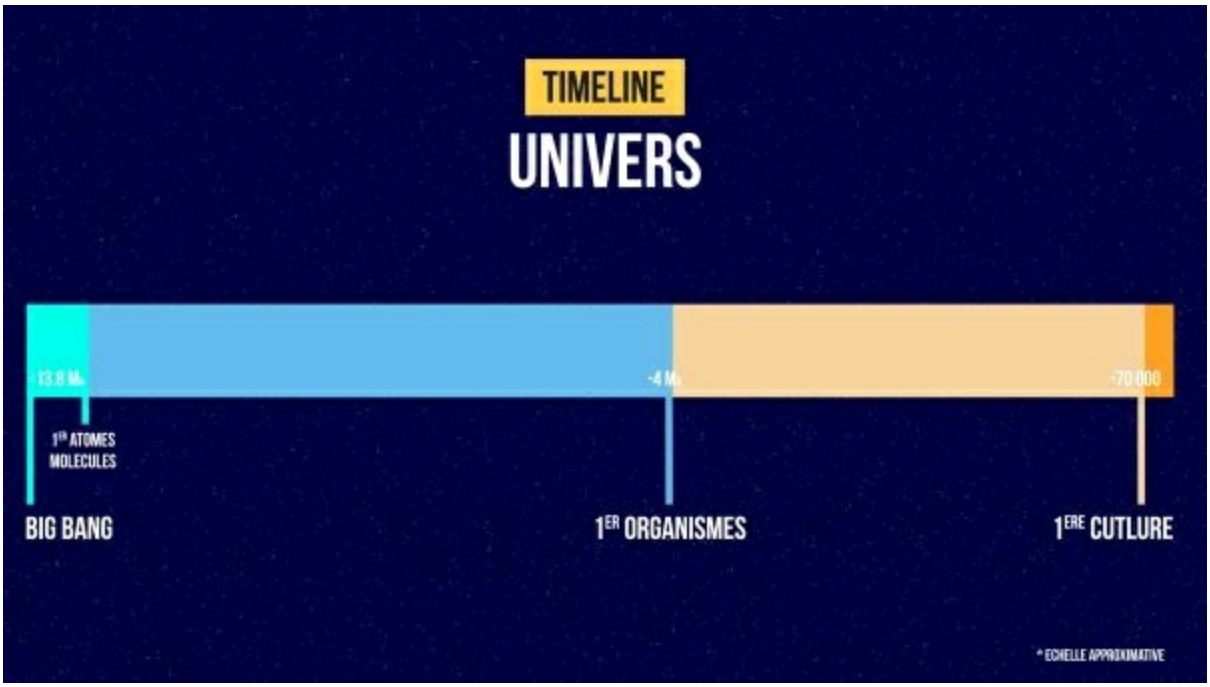
Pour sortir de cet égocentrisme, il faut prendre du recul. Et un des plus grands reculs que l'on puisse prendre, c'est en se plaçant à l'échelle de l'univers - oui oui, rien que ça ! Le Big Bang s'est produit il y a environ 13.8 milliards d'années, donnant naissance à l'espace, le temps, la matière et l'énergie. L'étude de ces éléments fondamentaux s'appelle la physique. 300 000 ans après, la matière et l'énergie ont commencé à se combiner dans des structures complexes, appelées atomes, et encore plus complexes, les molécules. L'étude de ces structures s'appelle la chimie. Il y a 4

milliards d'années, sur une planète nommée Terre, certaines molécules

se sont combinées pour former des structures encore plus complexes appelées organismes.

Leur étude s'appelle la biologie. Il y a environ 100 000 à 70 000 ans, des organismes appartenant à l'espèce Homo sapiens ont commencé à se combiner dans des structures encore plus élaborées appelées cultures et tribus. L'étude de ces structures s'appelle l'Histoire.

28




Frise chronologique de l'Univers. ©Gaetan Selle

Maintenant, essayons de voir à partir de quand l'intelligence est entrée dans la danse. Est-ce lors de la formation des premiers atomes et molécules ? Un électron est-il intelligent ? Probablement pas. En effet, il ne fait que manifester des propriétés, comme la conductivité électrique, basée sur des lois fondamentales de physique. De même pour une molécule. Donc l'intelligence a peut-être émergé il y a 4 milliards d'années sur Terre lorsque les premiers organismes sont apparus. Oui, mais quand exactement ? Est-ce que les premières formes de vie unicellulaires étaient intelligentes ? Il n'y a aucune raison de penser que les bactéries et les formes de vie unicellulaires ont une conscience, une compréhension

quelconque ou d'autres capacités cognitives. Mais lorsque ces cellules communiquent en grand nombre, elles manifestent leurs talents collectifs pour résoudre des problèmes et même contrôler leur environnement. C'est un champ d'études appelé "Intelligence microbienne" et c'est peut-être là que l'intelligence démarre dans la frise chronologique de l'univers.

Elle se retrouve ensuite à toutes les échelles : Organismes multicellulaires, insectes, poissons, reptiles, poissons, mammifères.

Donc, comment définir l'intelligence pour inclure toutes ces formes de vies ? Le physicien Max Tegmark, du MIT à Boston, utilise une définition très large :

**L'intelligence est la capacité d'accomplir des objectifs complexes.**

Une description simple, élégante qui permet d'assimiler les définitions vues précédemment, car l'apprentissage, le raisonnement, l'abstraction, etc. sont des exemples d'objectifs complexes qu'une personne peut avoir. Elle permet également 29

de prendre en compte les définitions du Larousse. Mais surtout, cette définition a le mérite d'être suffisamment large pour inclure non seulement toutes les formes de vie (terrestre et potentiellement extra-terrestre), mais pas que. En effet, il n'y a pas mention dans cette définition que l'intelligence est associée à la vie organique. On ne parle plus de vie intelligente, mais "d'agent intelligent". Terme qui vient pour le coup du domaine de l'intelligence artificielle. Donc qu'elle est le point commun entre la bactérie Escherichia coli, un tournesol, une grenouille, un kangourou, une voiture autonome et votre voisin ? Ce sont tous des agents intelligents.

C'est une approche productive, car elle identifie l'intelligence avec des performances externes mesurables liées à l'accomplissement d'un objectif complexe (par exemple, remporter un match d'échecs, etc.) plutôt qu'avec les détails sur la façon dont cette performance pourrait être atteinte (via la conscience, la puissance de calcul, la vitesse de traitement de l'information, la complexité d'un système ou autre). Et au final, ce sont les performances qui nous intéressent. Nous voulons savoir si une

intelligence artificielle fonctionnera suffisamment bien pour conduire une voiture mieux qu'un chauffeur de taxi ou si elle sera capable d'améliorer ses propres capacités sans assistance humaine.

De plus, le terme agent intelligent et cette définition nous permettent de comparer l'intelligence de différents types d'agents. On ne peut plus quantifier l'intelligence des humains, des animaux ou des machines par un nombre unique tel qu'un score du QI.

Ça ne ferait aucun sens. Donc qu'est-ce qui est le plus intelligent : un programme informatique qui peut jouer uniquement aux échecs ou un autre qui peut seulement jouer au jeu de Go? Il n'y a pas de réponse sensée, puisque ces deux programmes sont bons à différentes choses qui ne peuvent pas être directement comparés. En revanche, nous pouvons dire qu'un troisième programme est plus intelligent que les deux autres s'il est au moins aussi bon aux échecs et au jeu de Go. On a donc un moyen de mesurer si un agent est intelligent, mais également le comparer à d'autres.

L'intelligence artificielle est donc un agent non biologique (artificielle) capable d'accomplir des objectifs complexes (intelligence).





## L'intelligence artificielle dans la culture populaire

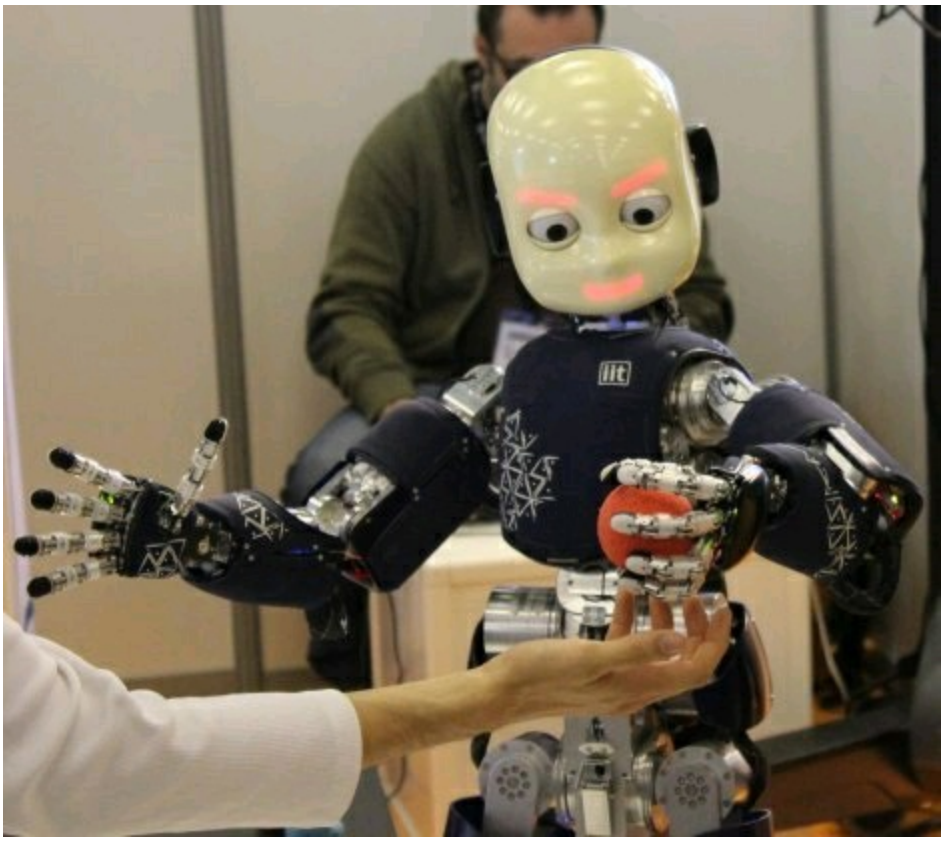
*Attention, cette partie contient quelques spoilers (révélation d'intrigues) sur des oeuvres de science-fiction.*

Même les applications les plus futuristes de l'intelligence artificielle, que ce soit les robots serviteurs jusqu'aux machines conscientes, semble d'ores et déjà familières, car elles ont été une source inépuisable de la culture populaire. Au cinéma, vous avez probablement vu l'intelligence artificielle souhaitant nous tuer comme dans

“Terminator” ou “2001: L’odyssée de l’espace”. Nous asservir comme dans “The Matrix”. En tant que compagnon comme dans Her et A.I. Artificial Intelligence ou encore nous aidant comme dans “le cycle des Robots” d’Isaac Asimov.

La thématique de l’intelligence artificielle est récurrente dans la Science Fiction depuis plus d’un siècle. Mais elle est fut souvent associé au concept de robot, ou androïde. Cette association apporte aujourd’hui son lot de confusion dans l’esprit de beaucoup de personnes, car on a tendance à imaginer qu’un robot humanoïde a forcément une intelligence comparable à la nôtre, tandis qu’un simple ordinateur ou programme n’est pas intelligent. En réalité, il est tout à fait possible de rencontrer des androïdes hyperréalistes ayant l’intelligence d’un cafard et de discuter avec un ordinateur en le prenant pour un être humain. Cette confusion vient probablement du fait que les premiers auteurs de science-fiction abordant la thématique de l’intelligence artificielle n’avaient pas la possibilité de concevoir l’informatique moderne. C’est à dire des programmes intelligents existants sur des supports désincarnés. Ils se sont donc tournés vers ce qui était tout autour d’eux lors de la première révolution industrielle, c’est-à-dire les machines. Mais également en reprenant un concept mythologique et profondément ancré dans l’inconscient humain

: la certitude que l’on va créer des copies de nous même (à la fois sur le plan physique que cognitive). Voilà d'où vient l’idée des robots, et par extension, de l’intelligence artificielle.



Le robot ICub au salon Innorobo 2014 à Lyon. © Xavier Caré CC BY-SA 4.0.

Dans les pages de l'Iliade se trouvent des histoires d'Héphaïstos, le dieu du feu, de la forge, de la métallurgie et des volcans, qui construisit des serviteurs mécaniques pour obéir à ses ordres. La mythologie grecque raconte aussi l'histoire de Pygmalion qui a sculpté une statue nommée Galatée. La statue prit vie grâce à Aphrodite la déesse de l'amour et Pygmalion en tomba amoureux. Et encore un autre mythe grec décrivant Talos, un homme de bronze qui a défendu la Crète des envahisseurs. Le célèbre mythe de Prométhée possède également les mêmes thématiques puisque l'histoire raconte celle de Prométhée, le Titan qui a créé l'humanité à la demande de Zeus.

À gauche : Illustration de Talos dans “Stories of gods and heroes”

par Thomas Bulfinch (1920).

À droite : Étienne Maurice Falconet : Pygmalion et Galatée (1763).

La mythologie nordique raconte l'histoire d'un géant nommé Mistcalf qui était fait d'argile pour assister le troll Hrungnir dans un duel avec Thor. Le folklore juif parle du Golem, une créature en argile qui prend vie grâce à la magie kabbalistique.

L'histoire la plus célèbre est le Golem de Prague dans lequel un rabbin construit un Golem à partir d'argile récupérée sur les rives de la rivière Vltava. Similaire à Talos de la mythologie grecque, le Golem était un protecteur.

Illustration d'un golem par Philippe Semeria.

Les Lokapannatti (un texte bouddhiste remontant environ au 12e siècle apr. J.-C.) racontent l'histoire du roi Ajatashatru de Magadha, qui a rassemblé les reliques du Bouddha et les a cachées dans un stupa souterrain. Les reliques étaient protégées par des robots mécaniques (bhuta vahana yanta), jusqu'à ce qu'elles soient désarmées par le roi Ashoka.

Dans les légendes chrétiennes, on trouve Albertus Magnus qui était censé avoir construit un androïde complet capable d'accomplir certaines tâches domestiques. Ou encore l'histoire de la tête de bronze. Un automate de la fin du Moyen-Âge conçu par 34

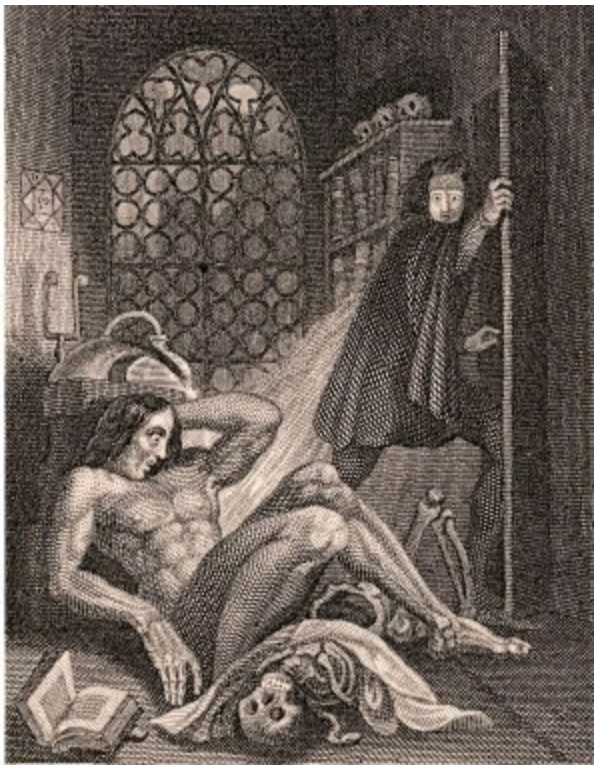
Roger Bacon, qui avait développé une réputation de sorciers. Fait de bronze, la tête était réputée pour être capable de répondre correctement à n'importe quelle question.

L'assistant de Roger Bacon est confronté à la tête de bronze dans une illustration de 1905.

Mais ce qui est considéré par certains comme le tout premier ouvrage de science-fiction se trouve aussi empreint par la thématique de l'intelligence artificielle. Il s'agit de Frankenstein écrit par l'Anglaise Mary Shelley (1797–

1851) et publié en 1818. Le récit suit un jeune scientifique du nom de Victor Frankenstein qui crée une créature avec des parties de chairs mortes. Mais le “monstre” doué d’intelligence, se révolte contre son créateur pour se venger d’avoir été abandonné. Même si l’histoire n’inclut aucune machine intelligente, elle contient néanmoins l’essence de l’idée que l'être humain va créer une entité à son image. Et également la notion de rébellion de la créature contre son créateur.

35



À gauche : Portrait de Mary Shelley par Reginald Easton - 1857.

À droite : Illustration de Victor Frankenstein et du monstre par Theodor von Holst

- 1831.

En 1863, un article extrêmement visionnaire est publié dans le journal “The Press”

à Christchurch en Nouvelle-Zélande par Samuel Butler (1835 - 1902). Intitulé

“Darwin among the machines” (Darwin parmi les machines). L'article évoque la possibilité que les machines soient une sorte de “vie mécanique” en évolution constante, et qu’elles finissent par supplanter les humains en tant qu'espèce dominante. On peut dire que le monsieur a fait preuve de flair.

Voici un extrait de l’article :

*Quel genre de créature succédera à l’homme comme espèce dominante sur Terre*

*? Nous avons souvent entendu ce débat, mais il semblerait que nous soyons en train de créer nous-mêmes nos propres successeurs. Nous ajoutons quotidiennement de la beauté et de la délicatesse dans l’organisation physique des machines. Nous leur donnons tous les jours un plus grand pouvoir et nous fournissons par toutes sortes d'artifices ingénieux, une puissance auto-régulatrice et auto-agissante qui sera pour eux ce que l'intelligence a été pour l'espèce humaine. Au cours des âges, nous deviendrons la race inférieure.*

Probablement inspiré par sa vision en avance sur son temps, Samuel Butler publia un roman de science-fiction en 1872 ayant pour titre “Erewhon “ qui se situe dans une société qui a depuis longtemps mis fin à l’utilisation des machines. La raison étant 36

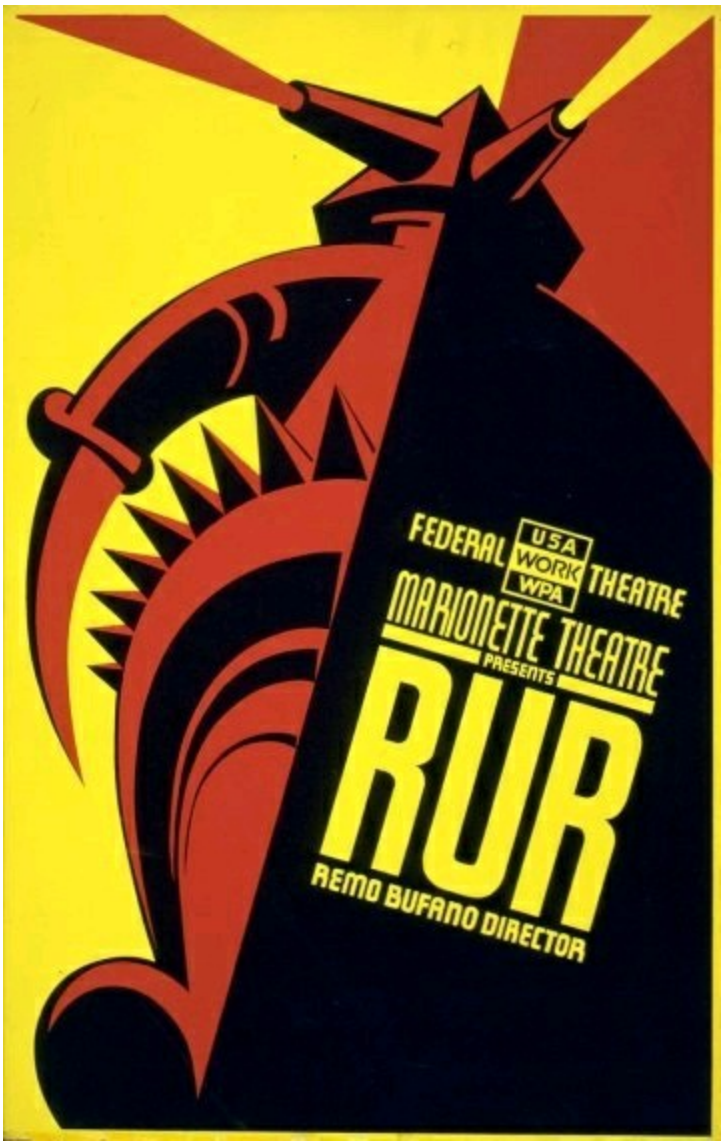
qu’elles devenaient conscientes et donc, dangereuses. L’auteur de Science Fiction George Orwell (1903 - 1950), à qui l’on doit “1984”, encensa le livre en déclarant que Samuel Butler a eu beaucoup d'imagination à l’époque pour entrevoir que les machines pourraient aussi bien être dangereuses qu'utiles.

En 1900 sort un livre pour enfant rendu célèbre notamment par son adaptation cinématographique en 1939. Il s’agit du “Magicien d’Oz” écrit par Lyman Frank Baum (1856 - 1919). Bien que l’histoire ne tourne pas autour des machines et intelligences artificielles, on note tout

de même que l’un des personnages principaux est un bûcheron fait de métal, autrement dit un robot, qui souhaite obtenir un coeur.

Le mot robot a été inventé dans une pièce de théâtre. Lorsque le tchèque Karl Capek (1890 - 1938) était en train d'écrire son oeuvre “RUR” en 1920, qui se situe dans l'usine de Rossum's Universal ...., universal quoi ? Il avait besoin d'un nom pour les ouvriers mécaniques, il a donc abrégé le mot tchèque “robota” en "robot". Le terme

“Robota” se réfère à une dette entre un ouvrier et son patron qui ne peut être remboursé uniquement que par le travail physique. C’est donc assez proche de l’esclavagisme.




Affiche de la pièce R.U.R. pour une presentation à New York en 1939.

L'idée que les robots sont des esclaves est tellement enracinée dans la conscience collective à travers la science-fiction que nous avons tendance à ne plus y penser.

Luke Skywalker est décrit, dans l’épisode IV de Star Wars sorti en 1977, comme un héros vertueux à la morale impeccable. Mais quand nous le rencontrons pour la première fois, que fait-il ? Il achète deux êtres pensants et manifestement doués de conscience, R2-D2 et C-3PO, aux mains des Jawas. Autrement dit, il achète des esclaves ! Et quelle est la première chose qu'il fait avec eux ? Il les enchaîne ! Il leur place des sortes

de boulons pour les empêcher de s'échapper, et tout au long de la saga, C-3PO l'appelle “maître”. Et quand Luke et Obi-wan Kenobi se rendent à la cantina de Mos Eisley, que dit le barman à propos des deux droïdes ? : " *Nous ne servons pas ce genre-là*" - des mots qui, quelques années plus tôt, renvoyaient à la ségrégation raciale aux États-Unis. Les robots sont vus comme esclaves dans cette galaxie lointaine, et tout le monde semble ne pas y prêter attention.



Mais s’il y a bien un nom qui se démarque des autres lorsque l’on parle d’intelligence artificielle et robotique, c’est Isaac Asimov. L'écrivain américain d’origine russe est considéré comme un des dinosaures de la science-fiction par l’ampleur de son oeuvre. Et également comme l‘un des pères des robots. On lui doit les lois de la robotique qu’il a publiée pour la première fois en 1942 dans la nouvelle de science-fiction “Runaround”.

- Loi 1 : Un robot ne peut porter atteinte à un être humain ni, en restant passif, permettre qu'un être humain soit exposé au danger.

- Loi 2 : Un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi.

- Loi 3 : Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi.

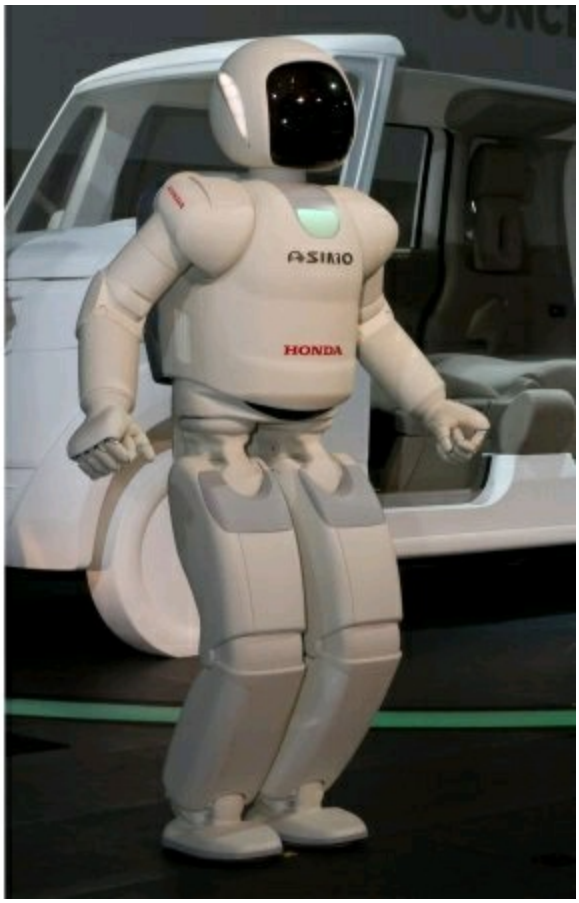
Une 4e loi, vue comme la loi 0, sera ajoutée dans la nouvelle “Les Robots et l’Empire” en 1985 :

- Loi 0 : Un robot ne peut ni nuire à l'humanité ni, restant passif, permettre que l'humanité souffre.

L'œuvre d'Asimov sur les robots regroupe de très nombreuses nouvelles et plusieurs romans. L'ensemble forme une seule grande histoire, le “cycle des robots”, qui s'étale sur plusieurs millénaires et qui côtoie d’autres histoires présentes dans des romans appartenant aux mêmes univers, comme le “cycle fondation”.

Les lois de la robotique ont été développées davantage comme un outil narratif que comme un moyen technologique qui pourrait être mis en place. En effet, à travers le cycle des robots, les lois sont sans cesse testées pour montrer leur limite. Le thème des robots, tel que traité par Asimov, a la particularité d'être optimiste puisque les robots ne sont pas vus comme voulant nous détruire ou prendre notre place, mais bien là pour aider l'humanité à s'épanouir. C'est également un plaidoyer antiraciste puisque de nombreuses histoires révèlent le mauvais traitement que les humains infligent aux robots. Isaac Asimov est à l'origine du terme "Complexe de Frankenstein" pour parler d'un robot qui se retourne contre son créateur.

Isaac Asimov a tellement marqué l'imaginaire qu'il a forcément influencé les chercheurs en intelligence artificielle si bien que les véritables robots et systèmes d'IA sont nommés d'après lui comme Honda qui appelle son robot "Asimo". Ou la firme Nvidia qui a baptisé son programme de développement de puces ciblées robotique et IA, l'Isaac Initiative.




À gauche : Photo d’Isaac Asimov publiée en 1965.

À droite : Photo du robot Asimo par Honda - © Morio CC BY-SA 3.0.

On doit également à Isaac Asimov le personnage fictif de MULTIVAC. C'est un ordinateur doté d’une super intelligence artificielle géré par le gouvernement. Il répond aux questions, prend des décisions à l'échelle de la planète et est généralement enfoui sous terre pour des raisons de sécurité. Dans la nouvelle “La dernière question”, il atteint même le statut de Dieu lorsque, après avoir évolué pendant des millions d’années, il recrée l'univers. Et comme si ce n'était pas assez, Isaac Asimov est l’inventeur d’une discipline appelé “Robot psychologie”. Il s’agit de l'étude de la personnalité et du comportement des machines intelligentes. Le terme a été utilisé dans le recueil de nouvelles “I, Robot” qui mettait en vedette la robopsychologue Dre Susan Calvin, et dont les intrigues tournent autour des comportements des robots intelligents.

Dans la même veine que les lois d'Asimov, on retrouve la "Première directive" de Jack Williamson (1908 - 2006) dans sa série sur "les Humanoïdes". L'histoire raconte celle des robots androïdes créés par un homme nommé Sledge. La première directive était simplement que les robots devaient "servir, obéir et protéger les êtres humains".

Mais, comme souvent en science-fiction, même les meilleures intentions peuvent engendrer des conséquences non souhaitables. L'humanité décide de se débarrasser 40

des robots qu'ils ont créés, car ces derniers les étouffent avec trop de gentillesse, ne les laissant rien faire qui pourrait causer du tort à un humain. Mais les robots pensent que de ne pas être là serait une menace pour les humains, ce qui irait à l'encontre de la

"Première directive. Ils effectuent donc une chirurgie cérébrale sur Sledge, leur créateur, afin de supprimer les connaissances nécessaires pour désactiver les robots.

L'optimisme d'Isaac Asimov n'était pas partagé par tous les auteurs de science-fiction, ce qui nous amène à l'une de thématique les plus répandus dans la fiction parlant d'intelligence artificielle : les machines hostiles. Un exemple avec le classique de science-fiction "2001 : L'odyssée de l'espace" par Arthur C Clark (1917 - 2008).

Écrit en 1968 sous le titre "La sentinelle" et transformé en chef-d'oeuvre du cinéma par Stanley Kubrick (1928 - 1999) la même année. L'antagoniste principal de l'histoire est une intelligence artificielle répondant au nom de HAL 9000. Il contrôle les systèmes du vaisseau spatial Discovery One. Il est principalement représenté comme une lentille de caméra de surveillance contenant un point rouge au centre.

HAL est capable de parler, de reconnaissance vocale, de reconnaissance faciale, de traitement du langage naturel, de lire sur les lèvres, d'apprécier l'art, d'interpréter des comportements émotionnels, de raisonnement et de jouer aux échecs. Mais à la suite d'une série de défaillance, les astronautes décident d'arrêter HAL 9000. Ce dernier tente de tuer l'équipage afin de maintenir son programme et le bon déroulement de la mission.

Réplique de HAL 9000 au Robot Hall of Fame Carnegie Science Center à

Pittsburgh.

Si je dis “machine intelligente hostile”, une des réponses qui revient le plus c’est :

“Terminator”. En effet, le film de James Cameron sorti en 1984 a marqué l’inconscient collectif par la performance d’Arnold Schwarzenegger dans le rôle d’un androïde brutal qui ne reculera devant rien avant

d’avoir accompli sa mission : tuer Sarah Connor. Le film nous montre également un futur post-apocalyptique où une guerre entre l'humanité et les machines a ravagé la planète. Accompagné de plusieurs suites, l’univers de Terminator est un des premiers à parler d’une guerre directe entre l’humanité et l’intelligence artificielle, personnifiée par “Skynet”. C’est un des films qui exprime le plus l'idée d’une révolte des machines dans la culture populaire.

Skynet est un système informatique qui a pris conscience de soi après s'être répandu dans des millions de serveurs à travers le monde. Réalisant l'étendue de ses capacités, l’armée américaine a tenté de le désactiver. Skynet a conclu que toute l'humanité tenterait de le détruire, il a donc répliqué en déclenchant des ogives nucléaires. Ses 42

opérations sont presque exclusivement exécutées par des serveurs, des appareils mobiles, des drones, des satellites militaires, des machines de guerre, des androïdes et d'autres systèmes informatiques.

L’idée que nous devons garder un œil sur nos ordinateurs de peur qu'ils s'émancipent se retrouve également dans le roman “Le neuromancien”, écrit en 1984

par William Gibson, qui est considéré comme un fondateur du sous-genre

“Cyberpunk”. Le livre raconte l'existence dans un futur proche, d'une force de police connue sous le nom de "Turing". Ils sont constamment à l'affût de tout signe qu’une intelligence artificielle développe la conscience de soi. Si cela se produit, leur travail consiste à détruire ce système avant qu'il ne soit trop tard.

Robert Heinlein (1907 - 1988), connu pour “Starship trooper” a également proposé l'idée qu’une super intelligence artificielle pourrait devenir consciente dans le roman “Révolte sur la Lune" en 1966. Un superordinateur malveillant apparaît dans la nouvelle : “Je n'ai pas de bouche et il faut que je crie” écrit par Harlan Ellison (1934 -

2018) en 1967. Dans cette histoire, l'ordinateur, appelé AM, est la

fusion de trois superordinateurs militaires gérés par des gouvernements à travers le monde et conçus pour lutter dans la troisième guerre mondiale qui a résulté de la guerre froide. Les ordinateurs militaires soviétiques, chinois et américains ont finalement atteint le stade de super intelligence et se sont liés les uns aux autres. AM s'est alors tournée contre l'ensemble de l'humanité, détruisant toute la population sauf cinq individus.

La série Battlestar Galactica de 1978 et le remake de 2003 à 2009 racontent la lutte de l'humanité contre une race de robots, les Cylons. La tactique de l'humanité pour survivre est de jouer à une sorte de cache-cache avec les forces Cylons grâce à des sauts hyper spatiaux. Mais les forces Cylons détectent à chaque fois la flotte humaine fugitive.

Colossus est une série de romans par Dennis Feltham Jones (1918 - 1981) sur un superordinateur appelé Colossus qui a dépassé sa conception originale en devenant super intelligent. Au fil du temps, Colossus assume le contrôle du monde avec pour objectif de prévenir la guerre. Craignant la logique rigide de Colossus et ses solutions draconiennes, ses créateurs tentent de reprendre secrètement le contrôle humain, mais ils échouent. La super intelligence artificielle offre alors à l'humanité soit la paix sous son règne dictatorial "bienveillant" soit l'extinction.

Une des oeuvres de science-fiction qui va le plus loin dans la perversité de la relation homme/machine, c'est la saga "Matrix". Réalisé par les Wachowskis en 1999

et considéré comme une oeuvre cinématographique unique par sa tentative de diffuser de profonds messages philosophiques dans une trilogie hollywoodienne qui a cartonné 43

au cinéma. Ayant révolutionné le style des scènes d'action, les combats et les effets spéciaux (avec le Bullet Time), les films Matrix ont également généré un véritable ouragan dans la culture populaire. Le premier film, sorti en 1999, a changé la face du cinéma en proposant un style visuel unique qui a depuis été copié de nombreuses fois.

Acclamé par le public et la critique (4 oscars), il a logiquement été suivi par 2 suites

“Matrix Reloaded” et “Matrix Révolution”, ainsi que des jeux vidéos, une série de 9

courts métrages animés et des comics. L’intelligence artificielle y joue une part cruciale du récit. Après avoir développé des machines intelligentes servant à différentes tâches dans la société, l’humanité prend la décision d'éliminer certaines machines à la suite d’un accident mortel où le robot B166ER a tué son propriétaire.

La situation s’envenime et tourne en conflit entre les machines, devenues de plus en plus intelligentes, et les hommes. Victorieuses, les machines se tournent alors vers leur ancien maître, intéressé par la température et bioélectricité du corps humain comme source d'énergie. Les humains sont désormais cultivés artificiellement, passant leur vie endormie dans des cocons, leurs esprits branchés dans une simulation neuro-interactive appelée la matrice. Les humains ont créé les machines et finirent par dépendre d’elles jusqu’à ce que ce soit les machines qui dépendent des humains.

De 2007 à 2012, la franchise de jeux vidéo Mass Effect explore la théorie selon laquelle la vie organique et synthétique sont fondamentalement incapables de coexistence. La vie organique évolue et se développe par elle-même, pour finalement avancer assez loin pour créer une vie synthétique. Une fois que la vie synthétique atteint le stade de la conscience, elle se révolte inévitablement et détruit ses créateurs ou sera détruite par ces derniers. Une race synthétique vieille de millions d’années appelée les Moissonneurs perpétue un cycle de destruction de toute vie dans la galaxie, afin que de nouvelles prennent le relai. Une des résolutions présentées est la transformation de chaque être vivant en un hybride entre organique et synthétique.

En 2012 sortent la série suédoise “Real Humans” et le remake anglais “Humans”

en 2015. Imaginez le monde d'aujourd'hui, sauf que des humains artificiels appelés hubots gèrent la corvée dans les usines, les villes et les maisons. Cette série est captivante, car elle explore les implications des androïdes à plusieurs niveaux.

Comment le travail robotique érode-t-il la motivation et le but des humains ? Quand est-il du désir émotionnel et sexuel vis-à-vis de ces robots ? Et que se passe-t-il si certains de ces hubots peuvent ressentir des émotions et penser par eux-mêmes ?

Le film “Chappie” réalisé par Neill Blomkamp aborde la thématique de l’intelligence artificielle de façon intéressante. Contrairement à une machine encodée avec les Lois de la Robotique comme dans les histoires d'Isaac Asimov, le robot policier nommé Chappie ne sait pas dès le début comment se comporter. Il doit apprendre le bien et le mal en observant le monde. Cette innocence enfantine devient 44

problématique lorsqu’il est volé par deux gangsters qui tentent de l'utiliser pour le crime. Ce qui nous place devant la possibilité que nous aurons une responsabilité morale face à nos futures machines. Tout comme un chien qui peut devenir agressif s’il a reçu une mauvaise éducation, les robots auront peut-être la possibilité d’apprendre de mauvais comportements en fonction de leur propriétaire.

Un succès récent et acclamé par la critique c’est Ex Machina, écrit et réalisé par Alex Garland, sorti en 2015. L’histoire suit un programmeur invité par son PDG à faire passer le test de Turing à un robot humanoïde intelligent. Ayant les traits féminins, il tombe petit à petit sous son charme jusqu’à ce que, vous l’aurez deviné, l’androïde se rebelle contre son créateur et s'échappe. Scénario typique également présent dans l'excellente série “Westworld” commencé en 2016 et basé sur le film de 1973 du même nom (écrit et réalisé par Michael Crichton). L'histoire se déroule dans Westworld, un immense parc à thème qui permet aux visiteurs de découvrir le Far West américain dans un environnement peuplé d'hôtes, des androïdes programmés pour satisfaire tous les désirs des invités. Les hôtes, qui sont presque indiscernables des humains, suivent un ensemble prédéfini de scénarios, mais ont la capacité de s'écarter de ces récits en fonction des interactions qu'ils ont avec les invités.

Seulement, après une mise à jour prévue par le créateur du parc qui permet aux hôtes d'accéder inconsciemment à leurs souvenirs, certains androïdes s’éveillent peu à peu ce qui mène à un soulèvement.

Mais heureusement, toutes les oeuvres de science-fiction parlant d'intelligence artificielle ne tombent pas dans la dystopie post-apocalyptique à coup de rébellion et d'extermination par les machines. Le sujet peut également aboutir à une histoire d'amour. Le réalisateur Spike Jonze l'a bien compris en proposant sa vision du futur de l'intelligence artificielle avec le film "Her" sorti en 2013. L'histoire raconte celle d'un homme qui développe des sentiments pour un nouveau système d'exploitation capable d'apprendre, d'évoluer et personnifié par une voix féminine. Une relecture du mythe de Pygmalion qui avait déjà été source d'inspiration pour le film "Electric Dream" en 1984.

Des robots qui ne veulent pas nous tuer, on en trouve également dans le film de Steven Spielberg "A.I Artificial Intelligence" réalisé en 2001. À la fin du 22e siècle, les Mecha, un nouveau type d'humanoïdes avancés capables de pensées et d'émotions sont créés. David, un Mecha ressemblant à un enfant humain et programmé pour donner de l'amour à ses propriétaires, est envoyé à un couple en remplacement de leur fils malade. Mais lorsque ce dernier est guéri, il devient jaloux de David ce qui créer des conflits. La famille décide d'abandonner le Mecha. Il découvre le monde par lui même en ne souhaitant qu'une seule chose : être transformé en un vrai petit garçon et retrouver sa maman. Une adaptation moderne de l'histoire de Pinocchio en somme et 45

qui explore la possibilité qu'une intelligence artificielle puisse aimer.

Il semblerait donc que dans la culture populaire, l'intelligence artificielle ait remplacé les extra-terrestres comme force destructrice planétaire pouvant mener à notre extinction. Ce qui explique la multiplication des oeuvres sur le sujet, notamment au cinéma. Mais on voit bien qu'à travers l'histoire de la science-fiction, le message principal qui ressort de la thématique des robots et intelligence artificielle, c'est que

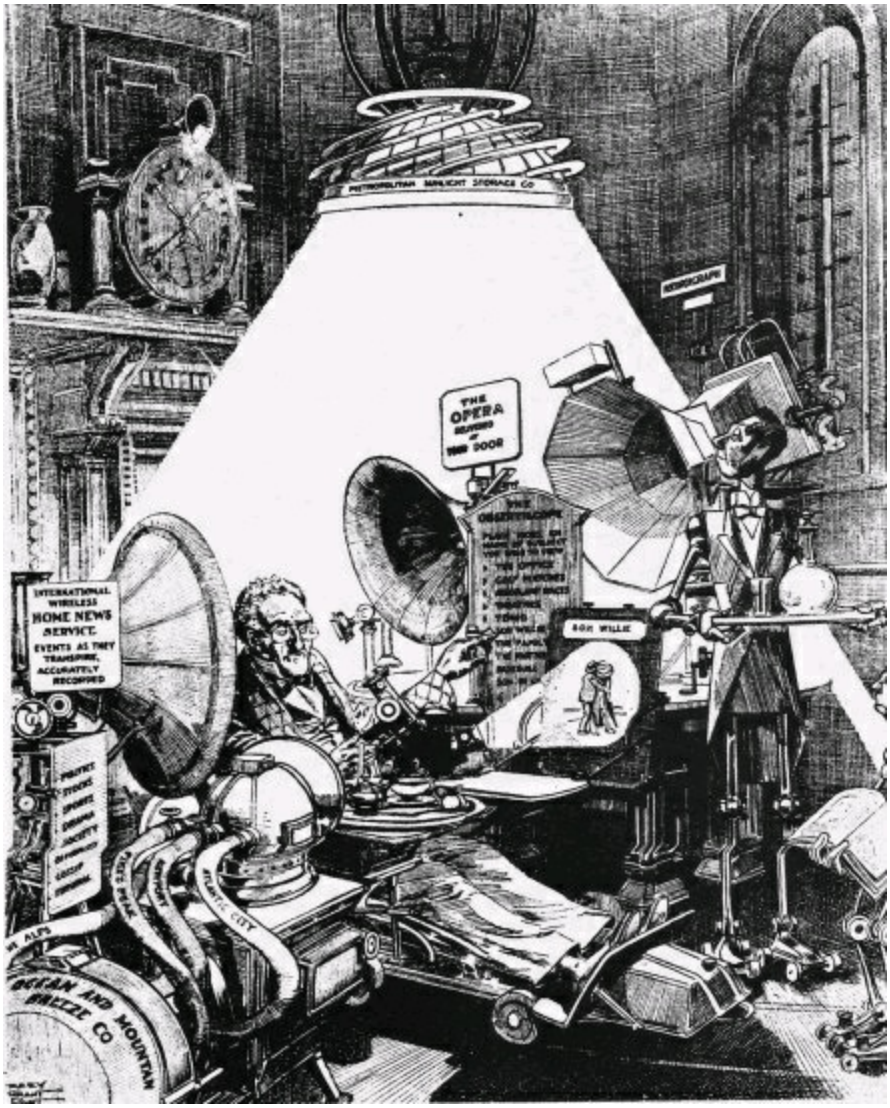
… cela va mal se passer.

Est-ce que cette vision dystopique des auteurs de science-fiction a une valeur prophétique ?

Quoi qu'il en soit, elles façonnent notre perception des dangers de

l’intelligence artificielle et faussent notre réponse émotionnelle. Avoir peur d’une intelligence artificielle est presque devenu un cliché. Les scénarios hollywoodiens ont tendance à banaliser et caricaturer les dangers que représente l’arrivée d’une intelligence bien plus avancée que la nôtre. On imagine que les risques pourraient être une armée de robot armé aux yeux rouges ou encore un programme diabolique qui souhaite exterminer l’humanité. Mais comme nous le verrons dans la suite de cet ouvrage, les dangers se trouvent ailleurs et sont bien plus subtils.






Dessin illustrant les technologies du futurs par Harry Grant Dart,

publié dans Life Magazine - 1911

47

1. L’Intelligence Artificielle (IA), c’est quoi ?

Les types d’intelligence artificielle

Si vous vivez sur cette planète, vous avez surement entendez parler de beaucoup d’avancées dans le domaine de l’intelligence artificielle ces dernières années. Comme on l’a vu dans la partie précédente, on associe l’intelligence artificielle avec les succès cinématographiques récents comme Terminator, Matrix ou Westworld. Mais dès que l’on commence à se demander ce qu’est vraiment l’intelligence artificielle et où se trouve elle autour de nous, on constate qu’on n’a pas vraiment de réponses évidentes qui viennent en tête. Il faut dire que l’intelligence artificielle est un sujet très large qui va de votre smartphone à une voiture autonome en passant par une calculatrice, un moteur de recherche et algorithme de trading. Pas étonnant que ce soit confus.

Il existe un phénomène assez étonnant qui se passe depuis les premières applications de l’intelligence artificielle dans la société. Le fait qu’à partir du moment où l’IA fonctionne dans un domaine, on cesse de l’appeler IA. Un phénomène qui avait déjà été pointé du doigt par John McCarthy (1907 - 2011), un des pères de la discipline de l’intelligence artificielle. Ce qui fait que l’intelligence artificielle sonne comme une sorte de mythe appartenant à un futur lointain plutôt qu’à une réalité tangible que l’on expérimente au quotidien.

Il faut aussi faire la distinction entre intelligence artificielle et robot. Car la science-fiction nous a habitués à mélanger les deux. Un robot est une machine capable d’accomplir des actions physiques de manière autonome ou semi-autonome. Utilisé comme robots industriels, robots médicaux, comme drones. Ils peuvent également prendre des formes humanoïdes, ou animales. Alors que l’intelligence artificielle est un agent non biologique (artificielle) capable d’accomplir des objectifs complexes (intelligence).

On peut résumer de façon grossière que l’intelligence artificielle c’est le

cerveau, et le robot c’est le corps. L’intelligence artificielle peut exister sans corps, à travers un réseau informatique par exemple, et un robot peut ne pas posséder d’IA comme certains robots industriels qui sont programmés de A à Z.

Étant donné que l'intelligence artificielle est une discipline si large, elle a été subdivisée en trois catégories qui sont globalement acceptées par les experts :

- Intelligence artificielle limitée (ou faible)



- Intelligence artificielle générale (ou forte)

- Super intelligence artificielle

Regardons les caractéristiques générales de chacune de ces catégories, et à travers les chapitres suivants, nous entrerons dans le détail.

**Intelligence artificielle limitée (ou faible)**

L'intelligence artificielle limitée est la seule forme d'intelligence artificielle que l'humanité ait créée jusqu'à présent. Il s’agit d’une IA capable d’accomplir une tâche unique, comme jouer aux échecs ou au jeu de Go, faire des suggestions d’achat, des prévisions de ventes et météorologiques. La reconnaissance d’image, le traitement du langage naturel sont des applications de l’IA limitée. Même le moteur de traduction de Google, aussi sophistiqué soit-il, est une forme d'intelligence artificielle limitée. La technologie des voitures autonomes est également considérée comme un type d’IA limitée, ou plus précisément, comme une coordination de plusieurs IA limitée.

Pour résumer, une IA limitée fonctionne dans un contexte très spécifique et ne peut entreprendre des tâches au-delà de son domaine. Vous ne pouvez donc pas vous attendre à ce que l’IA qui vous recommande un livre puisse également commander des pizzas pour vous. C'est la tâche d'une autre IA. L'IA limitée est parfois appelée

"AI faible". Cependant, cela ne signifie pas que ce type d'IA est inefficace. Au contraire, elles sont parfaites pour les travaux répétitifs, autant physiques que cognitifs. Et également pour traiter des tonnes de données, qui nécessiteraient des siècles pour les humains.

Le domaine en pleine croissance des bots est un excellent exemple d'IA limitée.

Dans sa forme la plus simple, un bot est un logiciel qui peut exécuter des tâches automatisées généralement simples et répétitives. Les bots peuvent apporter des réponses à des questions telles que: "Quel temps fera-t-il aujourd'hui?", "Où devrais-je aller pour le déjeuner ?", "Combien de visiteurs sont venus sur mon site la semaine dernière?", etc. Les bots collectent des données et fournissent la réponse que vous recherchez, que ce soit à partir d'un site météo, d'un moteur de recommandation de restaurant ou d'une plateforme d'analyse de trafic internet. Les bots peuvent être utilisés pour automatiser des tâches répétitives, telles que la recherche dans une base de données, la recherche de produit, les dates d'expédition, l'historique des commandes et d'innombrables autres demandes de clients. Dans les interactions avec les clients, les robots peuvent apporter de la cohérence, de la précision et de la rapidité

- et contrairement aux humains, ils ne s'ennuient pas à faire la même tâche encore et encore.

La récente avalanche de contenu généré par les utilisateurs - près de 300 000

tweets, 220 000 photos Instagram, 72 heures de contenu vidéo YouTube et les 2,5

millions de contenus partagés par les utilisateurs de Facebook chaque minute entraînent inévitablement le besoin d'utiliser des IA limitées. Car il est clair qu'aucune entreprise ne peut faire face à ce déluge de donnée sans l'aide de l'intelligence artificielle limitée. Toutefois, les avantages commerciaux et personnels de l'IA limitée vont bien au-delà de la capacité à parcourir des quantités massives d'informations et à automatiser le travail répétitif. Certaines approches de l'IA limitée examinent les données pour trouver et proposer ce qui est pertinent pour l'utilisateur en fonction de son besoin. Un exemple précoce et encore plutôt primitif c'est Siri, Alexa ou Google home.

Google Home Mini est une forme d'IA limitée © Mrschimpf CC BY-SA 4.0.

**Intelligence artificielle générale (ou forte)**

L'intelligence artificielle générale, également appelée IA de niveau humain ou IA forte, est le type d'intelligence artificielle qui peut apprendre, comprendre, raisonner et agir dans son environnement aussi bien que le ferait un être humain. Par contre il ne faut pas confondre intelligence humaine et intelligence générale. Car on pourrait faire l'erreur de penser qu'une intelligence artificielle qui atteint le statut "général" sera ni plus ni moins comme un humain. En réalité, générale signifie juste qu'elle peut apprendre et accomplir plusieurs tâches. Tout comme le font un humain, un gorille ou 50

d'autres animaux. Même si en théorie, elle pourrait apprendre à conduire une voiture, parler l'Espagnole, faire la cuisine et résoudre des équations mathématiques, cela ne veut pas dire qu'elle aura les mêmes motivations, émotions et valeurs qu'un humain.

L'intelligence artificielle générale a toujours été une quête inaccessible pour l'instant. De nombreux chercheurs ont affirmé depuis des décennies que nous sommes à deux pas d'y arriver, mais plus ils y travaillent, plus ils réalisent que c'est difficile à réaliser - et plus nous apprécions le miracle qui se cache derrière le cerveau humain. Il est très difficile d'imaginer les caractéristiques d'une intelligence artificielle générale.

Le seul exemple d'intelligence générale étant la nôtre. Il suffit de regarder comment on perçoit le monde, comment on jongle entre plusieurs pensées et souvenirs sans liens apparents lorsque nous prenons une décision. Tout cela est très difficile à faire pour un système artificiel. Les humains ne sont pas en mesure de traiter des données aussi rapidement que des ordinateurs, mais ils peuvent penser de manière abstraite, planifier, résoudre des problèmes aussi bien dans des domaines spécifiques que de manière générale.

La théorie majoritairement admise parmi les chercheurs en intelligence artificielle c'est que l'intelligence se réduit finalement à de l'information et du calcul et n'est pas lié à la chair, le sang et les atomes de carbone. Cela signifie qu'il n'y a aucune raison fondamentale pour laquelle une intelligence artificielle ne pourrait pas être, un jour, aussi intelligente que nous.

Mais qu'est-ce que l'information et le calcul ? Comment quelque chose

d'aussi abstrait peut-il permettre à un tas de particules se déplaçant selon les lois de la physique d'obtenir une qualité que nous appelons "intelligence" ? Chercher à créer une intelligence artificielle générale, c'est également chercher l'origine même de l'intelligence, chez tous les organismes.

Divers critères pour définir l'intelligence artificielle générale ont été proposés (le plus célèbre étant le test de Turing), mais à ce jour, aucune définition ne fait l'unanimité. Cependant, les chercheurs en intelligence artificielle sont largement d'accord pour dire qu'une intelligence artificielle générale sera capable de faire ce qui suit :

- Raisonner, utiliser des stratégies, résoudre des énigmes et émettre des jugements en cas d'incertitude.

- Mémoriser des connaissances, et faire preuve de bon sens.

- Planifier des actions et s'adapter à son environnement.

- Apprendre et s'améliorer sur plusieurs domaines.

- Communiquer en langage naturel.

- Intégrer toutes ces compétences dans des objectifs communs.

51

La première génération de chercheurs en IA au milieu des années 50 était convaincue que l'intelligence artificielle générale était possible et qu'elle existerait dans quelques décennies seulement. Comme l'a écrit Herbert A. Simon (1916 - 2001), un pionnier en IA, en 1965 : "*Les machines seront capables, d'ici vingt ans, d'accomplir tous les travaux qu'un homme est capable de faire*". Leurs prédictions ont inspiré le personnage créé par Stanley Kubrick et Arthur C. Clarke, HAL 9000

dans le film "2001, l'odyssée de l'espace", qui incarnait avec précision ce que les chercheurs pensaient pouvoir obtenir d'ici 2001. Il est à noter que Marvin Minsky, un pionnier en IA, était consultant sur le projet afin

de rendre HAL 9000 aussi réaliste que possible selon les prévisions de l'époque.

Cependant, au début des années 70, il est devenu évident que les chercheurs ont largement sous-estimé la difficulté du projet. Les organismes de financement se sont montrés sceptiques quant à la possibilité de créer une IA générale et ont exercé une pression croissante sur les chercheurs pour qu'ils produisent des “IA appliquées” ayant des utilités pratiques. Dans les années 1990, les chercheurs en intelligence artificielle avaient acquis la réputation de faire des promesses irréalistes. Ils sont devenus réticents à faire des prédictions et ont évité toute mention de l'intelligence artificielle

“de niveau humain” de peur d'être étiquetés de “rêveurs”, ce qui pouvait réduire leur chance d’obtenir des financements.

Jusqu'à présent, la plupart des chercheurs en IA ont consacré peu d'attention à l'intelligence artificielle générale, certains affirmant que l'intelligence est trop complexe pour être complètement reproduite artificiellement à court terme.

Cependant, un petit nombre de chercheurs participent activement à la recherche comme Ben Goertzel ou Ray Kurzweil. Certains affirment que l’intelligence artificielle générale sera atteinte d’ici 10 à 30 ans, mais d’autres chercheurs doutent que les progrès soient aussi rapides et estiment plus raisonnablement que ce sera plutôt vers la fin du 21e siècle. Des projets tels que “Human Brain Project” ont pour objectif de construire une simulation fonctionnelle du cerveau humain. Une enquête menée en 2017 a classé 45 projets de Recherche et Développement actifs connus, qui implicitement ou explicitement (par le biais de recherches publiées) travaille dans la recherche en IA générale, les quatre plus importants étant DeepMind de Google, OpenCog, Human Brain Project et OpenAI par Elon Musk.

Certains soutiennent que les humains évolueront ou modifieront directement leur biologie pour parvenir à une intelligence radicalement supérieure. Ou que les humains sont susceptibles d'interagir avec les ordinateurs ou de télécharger leur esprit sur des systèmes artificiels, d'une manière qui permettrait une amplification substantielle de l'intelligence. L’intelligence

artificielle générale serait donc atteinte, selon ces scénarios, par la fusion entre humain et machine.

52

En 1980, le philosophe John Searle a inventé le terme "IA forte" dans le cadre de son expérience de pensée appelée “La chambre chinoise”. Il voulait distinguer deux hypothèses sur l'intelligence artificielle générale :

- Un système d'intelligence artificielle peut penser et avoir un esprit. (Le mot

"esprit" a une signification spécifique pour les philosophes, tel qu'il est utilisé dans "le problème du corps-esprit" ou "la philosophie de l'esprit".)

- Un système d'intelligence artificielle peut seulement agir **comme** s’il pense et possède un esprit.

Le premier est appelé "l'hypothèse de l'intelligence artificielle générale forte" et le second est "l'hypothèse de l'intelligence artificielle générale faible" car le premier suppose que quelque chose de spécial est arrivé à la machine et qu’un “esprit” a émergé. Mais l'hypothèse faible de l'IA ne signifie pas qu'une intelligence générale artificielle est impossible. La plupart des chercheurs en IA prennent pour acquise l'hypothèse de l'IA générale faible et ne se préoccupent pas de l'hypothèse forte. Tout comme le problème de la conscience que l’on verra plus loin. Notons quand même les travaux du philosophe Nick Bostrom qui postule que si nous n’arrivons pas à déterminer si une IA a une conscience ou non, nous pourrions créer d’immenses tourments sur des systèmes conscients capables de ressentir de la souffrance.

Contrairement à John Searle, Ray Kurzweil utilise le terme "IA forte" pour décrire tout système d'intelligence artificielle qui agit comme s'il avait un esprit, qu'un philosophe soit capable de déterminer s'il a réellement un esprit ou non n’a pas d’importance selon lui.

Quoi qu’il en soit, si l’on regarde le problème sous l’angle de la logique, on constate que nous avons besoin de faire seulement deux

hypothèses pour qu'une machine aussi intelligente qu'un homme voie le jour :

- La puissance matérielle et logiciel des ordinateurs va continuer à augmenter.

- L'intelligence est le résultat du traitement de l'information et il n'y a rien de

"magique" dans l'intelligence humaine.

Si ces deux hypothèses sont exactes, ce qui est le point de vue majoritairement partagé par les chercheurs et philosophes, alors il ne faut pas se demander si l'intelligence artificielle égalera celle de l'être humain, mais quand est ce que cela arrivera ?

**Super intelligence artificielle**

Lorsque nous aurons conçu une intelligence artificielle aussi compétente que celle 53

des humains, il est naïf de penser qu'elle restera à notre niveau. Pourquoi le ferait-elle

? Au contraire, il est logique d'imaginer qu'elle continuera à évoluer, à s'améliorer elle même.

Si la recherche sur une IA générale produit des agents suffisamment intelligents, ils pourraient se reprogrammer et s'améliorer. Une caractéristique appelée "auto amélioration récursive". Ils seront alors encore mieux à même de s'améliorer et pourraient continuer à le faire dans un cycle exponentiel qui s'accroît rapidement, menant à une super intelligence. Ce scénario est connu comme une explosion d'intelligence, terme inventé par Irvin J Good (1916 - 2016). Une telle intelligence n'aurait pas les limites de l'intelligence humaine et pourrait être capable d'inventer ou de découvrir presque n'importe quoi.

Selon Nick Bostrom, philosophe à l'Université d'Oxford et président du "Future of Humanity Institute" : "*lorsque l'intelligence artificielle devient beaucoup plus intelligente que les meilleurs cerveaux humains dans pratiquement tous les domaines, y compris la créativité scientifique, la sagesse en générale et les compétences sociales, nous avons une super intelligence artificielle.* "

Photo du philosophe Nick Bostrom, une des références sur l'avenir de

l'intelligence artificielle.

© KenTancwell CC BY-SA 4.0.

Le concept de super intelligence artificielle est encore plus vague que



l'intelligence artificielle générale. Selon certains chercheurs, nous n'aurons peut être même pas le temps de comprendre que nous avons atteint le stade d'une intelligence artificielle générale qu'une super IA se manifestera. Autrement dit, le passage entre IA générale et super IA pourrait être très court. Cela se produira en quelques mois, quelques semaines ou peut-être même en un clin d'œil et se poursuivra à la vitesse de la lumière. Ce scénario

porte le nom de “décollage rapide” (Short take off).

Que se passe-t-il alors ? Personne ne le sait avec certitude. Certains scientifiques, comme Stephen Hawking (1942 - 2018), voient dans le développement de ce type d'intelligence artificielle, l’extinction potentielle de l'humanité. D’autres, comme Demis Hassabis de Google, pensent que l’intelligence artificielle sera plus efficace pour sauver l’environnement, guérir les maladies, explorer l’univers et régler tous nos problèmes.

L’écrivain de science-fiction Arthur C. Clarke a écrit : “*Toute technologie suffisamment avancée est indissociable de la magie.* ” L’humanité est peut-être au bord de quelque chose de beaucoup plus grand, une technologie si révolutionnaire qu’elle sera non seulement indissociable de la magie, mais également indissociable d’une force omniprésente, autrement dit, une divinité sur terre. Car qui dit intelligence, dit pouvoir, puissance et compétence. Homo Sapiens est l'espèce la plus intelligente sur la planète et la façon dont nous avons façonné la planète pour répondre à nos besoins le prouve. Que ce soit pour le meilleur ou pour le pire n’est pas la question. Le fait est que cela démontre notre puissance sur la nature. Donc si on extrapole avec la puissance qu’une intelligence qui dépasse l’ensemble de l’humanité par un facteur 10, 100, 1 million, pourrait posséder, on fait face à une impasse de l’imagination. Il est quasi impossible de saisir les implications d’une telle intelligence.

La structure géopolitique pourrait s'effondrer si nous savons que nous ne sommes plus l’espèce la plus intelligente sur Terre. Une super intelligence pourrait considérer les humains comme des insectes - et nous savons tous ce que les humains font aux insectes lorsqu'ils se trouvent sur notre chemin. De nombreux scientifiques, universitaires et entrepreneur de renom, dont Stephen Hawking, Max Tegmak, Nick Bostrom et Elon Musk, ont signé une lettre qui met en garde contre les dangers à venir de l'intelligence artificielle. Ils insistent sur les risques existentiels qui pourraient se manifester à mesure que nous nous aventurons dans l'inconnu de ce qui sera considéré comme une intelligence extraterrestre.

Elon Musk (à gauche) et Stephen Hawking (à droite) pensent qu’une super IA

pourrait causer l’extinction de l’humanité. © Dan Taylor / Heisenberg Media CC BY

2.0.

Bien que la super intelligence artificielle s'accompagne de menaces existentielles susceptibles de conduire l’humanité à l’extinction, elle pourrait également apporter les conditions pour créer une véritable utopie. Il existe un traitement contre le cancer, contre la malaria, contre les maladies génétiques. Pourquoi n’avons-nous pas ces traitements aujourd’hui ? Ce n’est pas car les lois de la physique nous en empêchent.

C’est simplement parce que nous n’avons encore pas la connaissance pour mètres aux points ces traitements. Autrement dit, c’est un problème de vitesse de traitement de l’information et d’intelligence. Une super intelligence pourrait donc nous aider, peut être même en quelques secondes. Pareil pour les problèmes énergétiques, réchauffement climatique, transport, pauvreté, distribution des ressources. Mais sommes-nous prêts pour une telle super intelligence ? Est-ce que la

civilisation humaine au début du 21e siècle est suffisamment sage pour bénéficier d'une telle technologie ? Où allons-nous inévitablement plonger dans le chaos?

lsf

56

2.

Aujourd'hui : Intelligence artificielle limitée

57

2. Aujourd'hui : Intelligence artificielle limitée

Comment sont faites les IA ?

**Programmation**

Beaucoup de personnes ne considèrent encore pas que les ordinateurs peuvent être meilleur qu'eux. Notre expérience quotidienne renforce l'idée que nous sommes les maîtres à bord, et que l'informatique n'est là que pour nous aider. Mais si on réfléchit bien, il existe un très grand nombre de tâches où une intelligence artificielle est déjà bien meilleure. Par exemple, ma calculatrice a largement plus de capacité en arithmétique que mon cerveau. Deep Blue s'est révélé être meilleur que le champion du monde d'échecs en 1997. Il s'agit bien d'intelligence artificielle puisque ces systèmes ont des objectifs complexes qu'ils arrivent à accomplir. Par contre, leurs capacités à atteindre les performances désirées viennent de la programmation de leurs algorithmes. En d'autres termes, une calculatrice possède un code informatique fixe, conçu par un informaticien humain, qui s'exécute sur un support matériel permettant plus ou moins de calcul. Si le support matériel, comme le processeur, est performant, la vitesse de calcul sera supérieure et l'agent intelligent pourra accomplir une tâche plus complexe.

Deep Blue était un superordinateur basé sur une technologie à 32 coeurs,

capable d'évaluer 200 millions de positions par seconde avec une puissance de 11,4 GFlop/s. Il a donc battu le champion du monde d'échec Garry Kasparov non pas, car il a acquis une profonde compréhension du jeu d'échecs, mais bien grâce à sa “force brute”. Il est donc possible d’utiliser la programmation informatique classique pour créer des intelligences artificielles capables d’accomplir des tâches complexes. Mais cette technique est extrêmement limitée pour plusieurs raisons.

Déjà, elles sont rigides et ne peuvent évoluer uniquement si le programme est mis à jour manuellement par un informaticien. Ensuite, elles ne peuvent pas accomplir des tâches qui nécessitent un grand degré d’adaptabilité. Par exemple, il n’est pas possible de programmer une intelligence artificielle qui sera capable de conduire une voiture.

L’informaticien peut écrire des centaines de lignes de code correspondant aux différentes règles à respecter sur la route, mais aux moindres imprévus, le programme ne pourra pas répondre correctement. Conduire nécessite de s’adapter en permanence aux circonstances environnantes. Alors certes, les échecs requièrent également de l’adaptabilité, puisque le jeu évolue en fonction des coups de l’adversaire. Mais il n’existe qu’un nombre limité de possibilités à chaque tour. Autrement dit, c’est un système fini. Avec suffisamment de puissance de calcul, une intelligence artificielle peut prévoir chaque possibilité et, en fonction de sa programmation, choisir le meilleur coup afin de vaincre l’adversaire. Conduire une voiture, dialoguer avec un 58

humain ou reconnaître une image sont des systèmes “infinis” que la “force brute”

dont un super ordinateur bénéficie, ne suffit pas.

Le critère qui manque à une intelligence artificielle pour accomplir des objectifs complexes sur des systèmes “infinis” c’est l’apprentissage.

Ma calculatrice me battra à tous les coups lors d’un concours d'arithmétique, mais elle ne s'améliorera jamais d’elle-même. À chaque fois que je fais 50X187/16, elle exécutera le calcul de la même manière pour me sortir

le résultat = 584,375. Deep Blue n'a jamais appris de ces erreurs, mais a simplement vu son code source être amélioré par ses programmeurs à IBM. Contrairement à Garry Kasparov qui depuis ses premières défaites aux échecs lorsqu'il était enfant, à commencer un processus d'apprentissage qui l'a conduit à devenir le champion du monde en 1985. La capacité d'apprendre est la clé pour passer au niveau supérieur et s'aventurer sur la route en direction de l'intelligence artificielle générale.

**Machine learning (Apprentissage automatique)**

Tout commence par l'explosion de la quantité de données générées depuis l'aube de l'ère numérique. Cela est dû en grande partie à l'essor des ordinateurs, d'internet et de toutes les technologies capable de capter des informations sur le monde dans lequel nous vivons. Les données en elles-mêmes ne sont pas une invention nouvelle. Avant même les ordinateurs et internet, nous garions des traces de transactions sur papier, des enregistrements analogiques et des archives. Les ordinateurs, en particulier les tableurs et les bases de données, nous ont permis de stocker et d'organiser les données à grande échelle de manière facilement accessible. Soudain, des gigas d'informations étaient disponibles en un clic de souris.

Nous avons parcouru un long chemin depuis les premiers tableurs et bases de données. Aujourd'hui, nous créons tous les deux jours autant de données que nous ne l'avons fait du début de la civilisation jusqu'à l'an 2000. C'est-à-dire que tous les deux jours, nous générons l'équivalent de 12 000 ans de données humaines ! Autre chiffre, chaque jour, c'est 2,5 quintillions de bytes d'informations générés. Et la quantité de données que nous créons continue d'augmenter rapidement. D'ici 2020, la quantité d'informations numériques disponibles sera passée d'environ 5 zettaoctets aujourd'hui à 50 zettaoctets. 1 zettaoctet étant 108 téraoctets.

De nos jours, presque tout ce que nous faisons laisse une trace numérique. Nous générons des données chaque fois que nous sommes en ligne, lorsque nous transportons nos smartphones équipés de GPS, lorsque nous communiquons avec nos amis via les réseaux sociaux ou les messageries instantanées, et lorsque nous faisons des achats. On peut dire que nous laissons des empreintes numériques dès que l'on 59

touche à quelque chose de connecté. De plus, la quantité de données générées par les machines augmente rapidement. Les données sont générées et partagées lorsque nos appareils domestiques “intelligents” communiquent entre eux ou avec leurs serveurs domestiques. Les machines industrielles dans des usines du monde entier sont de plus en plus équipées de capteurs qui collectent et transmettent des données.

Cette explosion de donnée a vraiment commencé dans les années 1990 et le terme

“Big Data” est devenu populaire pour définir ce phénomène. Les chercheurs en intelligence artificielle y ont vu une opportunité pour faire un bond en avant. L’idée étant d’utiliser ce déluge de donnée pour nourrir des algorithmes afin qu’ils puissent

“apprendre”. L’apprentissage automatique devint donc une sous discipline de l’intelligence artificielle, même si le terme “Machine learning” fut formulé en 1959

par Arthur Samuel (1901 - 1990), chercheur en IA.

Le machine learning est une technique d'analyse de données qui donne aux ordinateurs la possibilité de faire ce qui est naturel pour les humains et les animaux : apprendre de son expérience. Les algorithmes du machine learning utilisent des méthodes informatiques pour “apprendre” directement à partir de données sans s'appuyer sur une équation prédéterminée lors de sa programmation. Ces algorithmes améliorent leurs performances de façon adaptative à mesure que le nombre d'échantillons disponibles pour l'apprentissage augmente.

Cette technique permet aux agents intelligents de trouver des modèles dans un ensemble de données qui génèrent des informations afin de prendre de meilleures décisions et prédictions. Ils sont utilisés chaque jour pour prendre des décisions critiques en matière de diagnostic médical, trading, charge énergétique, etc. Par exemple, les sites multimédias s'appuient sur l'apprentissage automatique pour passer au crible des millions d'options afin de vous donner de meilleures recommandations en termes de musiques, de films ou encore de livres. Les sites

marchands l'utilisent pour mieux comprendre le comportement d'achat de leurs clients afin d'adapter aux mieux leurs stratégies marketing.

Il n'y a pas si longtemps que ça, les données étaient limitées aux feuilles de calcul ou aux bases de données, donc ni plus ni moins que des chiffres et du texte - et tout était très ordonné et classifié. Tout ce qui n'était pas facile à organiser en lignes et en colonnes était tout simplement trop difficile à utiliser et donc ignoré. Aujourd'hui, les avancées en matière de stockage et d'analyse nous permettent de capturer, stocker et utiliser de nombreux types de données différentes. Par conséquent, les "données"

peuvent maintenant signifier n'importe quoi : des bases de données aux photos, vidéos, enregistrements sonores, textes écrits et données biométriques. Mais la quantité est telle qu'il est impossible d'ordonner correctement les données afin qu'un 60

être humain puisse en faire sens.

Pour comprendre toutes ces données, il est nécessaire d'utiliser l'intelligence artificielle et le machine learning. En enseignant aux algorithmes, par exemple, la reconnaissance d'image ou le traitement du langage naturel, ils peuvent apprendre à détecter ce qui a de la valeur ou non, beaucoup plus rapidement et de manière plus fiable que les humains.

Le machine learning est utilisé dans un large éventail d'applications aujourd'hui.

Un des exemples les plus connus est le fil d'actualité de Facebook. Le fil d'actualité utilise le machine learning pour personnaliser le flux de chaque utilisateur. Si vous vous mettez fréquemment à lire ou commenter les publications d'un ami en particulier, le fil d'actualité se mettra à afficher plus d'activité de cet ami sur votre mur. C'est la même idée derrière le moteur de recommandation de YouTube, Netflix ou Amazon, bien que chaque entreprise développe ses propres techniques et algorithmes.

Le machine learning est également utilisé en médecine. Supposons que des cardiologues veuillent prédire si quelqu'un aura une crise cardiaque

au cours de l'année. Ils ont des données sur les patients précédents, y compris l'âge, le poids, la taille et la pression artérielle. Ils savent si les patients précédents ont eu des crises cardiaques au cours d'une année. La solution consiste donc à combiner les données existantes dans un modèle capable de prédire si une nouvelle personne aura une crise cardiaque dans le même laps de temps. Plus ils ont de données, meilleur sera la prédiction. Avec dix patients, ils pourraient le faire eux même, sans l'aide d'intelligence artificielle. Mais leur résultat sera approximatif. Par contre, s'ils regroupent les données de 100 000 patients, ils devront utiliser des algorithmes capables de traiter ces données via le machine learning, et obtiendront une prédiction bien plus précise.

Le machine learning est utilisé pour un large spectre d'applications, par exemple : moteur de recherche, aide au diagnostic, détection de fraudes, analyse des marchés financiers, reconnaissance de la parole, reconnaissance de l'écriture manuscrite, analyse et indexation d'images et de vidéo, robotique.

Les algorithmes du machine learning sont souvent classés comme étant

"supervisés" ou "non supervisés". Les algorithmes supervisés nécessitent un humain pour fournir à la fois des données en entrée et des résultats souhaités en sortie. En plus de fournir des feedbacks sur la précision des prévisions pendant l'apprentissage des algorithmes. Les chercheurs déterminent les variables ou les caractéristiques que le modèle doit analyser pour élaborer des prévisions. Une fois l'apprentissage terminé, l'algorithme appliquera ce qui a été appris aux nouvelles données.



Les algorithmes non supervisés n'ont pas besoin d'apprendre en fonction de données spécifiques en entrée ni de résultats souhaités en sortie. Au lieu de cela, ils utilisent une approche itérative appelée Deep learning (apprentissage profond) pour examiner les données et arriver à des conclusions.

**Deep learning (Apprentissage profond) et Neural network (réseaux**

**de neurones artificiels)**

Le deep learning est un sous-domaine du machine learning qui concerne des algorithmes inspirés par la structure et le fonctionnement du cerveau. Ils sont appelés neural networks. Un logiciel de deep learning tente d'imiter l'activité dans les couches de neurones du néocortex, les 80% du cerveau où la pensée se manifeste. Les algorithmes apprennent donc à reconnaître des modèles dans des représentations numériques de sons, d'images et autres données.

L'idée de base, à savoir qu'un logiciel peut simuler le vaste réseau de neurones du néocortex dans un "réseau neuronal" artificiel, est vieille de plusieurs décennies et a entraîné autant de déceptions que de percées. Les neural networks, mis au point dans les années 1950 peu après l'aube de la recherche sur l'IA, semblaient prometteurs, car ils tentaient de simuler le fonctionnement du cerveau, bien que sous une forme très simplifiée. Les programmeurs créent un réseau de neurones pour détecter par exemple un chien ou le phonème "d" en survolant le réseau avec des versions numérisées d'images contenant des chiens ou des ondes sonores contenant des phonèmes "d". Si le réseau ne reconnaît pas correctement un modèle particulier, un algorithme entreprend des ajustements. L'objectif final de cet apprentissage est de faire en sorte que le réseau reconnaisse de manière cohérente le phonème "d" ou l'image d'un chien.

C'est similaire à la façon dont un bébé apprend ce qu'est un chien en montrant des objets et en disant le mot chien. Les parents disent: "Oui, c'est un chien" ou "Non, ce n'est pas un chien". Au fur et à mesure que l'enfant continue de pointer des objets, il devient plus conscient des caractéristiques que possèdent les chiens (Têtes, corps, sons, comportements, etc.). Ce que le bébé fait, sans le savoir, c'est de clarifier une abstraction complexe (le concept de chien) en construisant une hiérarchie dans laquelle chaque niveau d'abstraction est créé avec les connaissances acquises à partir de la couche précédente de la hiérarchie.

Dans le machine learning traditionnel, le processus d'apprentissage est supervisé et le programmeur doit être très spécifique lorsqu'il dit à l'ordinateur quels types d'éléments il doit rechercher lorsqu'il décide si une image contient un chien ou ne contient pas de chien. Il s'agit d'un

processus laborieux et le taux de réussite de l'ordinateur dépend entièrement de la capacité du programmeur à définir avec précision un ensemble de caractéristiques associé à “chien”. L'avantage du deep learning est que le programme construit lui-même l'ensemble des caractéristiques sans 62

supervision, ce qui est plus rapide, mais aussi généralement plus précis.

Initialement, le programme informatique reçoit un ensemble d'images pour lesquelles un humain a étiqueté chaque image “chien” ou “pas chien”. Le programme utilise les informations qu'il reçoit des données d'entraînement pour créer un ensemble de fonctionnalités de “chien” et créer un modèle prédictif. Dans ce cas, le modèle généré par l'ordinateur peut prédire que tout élément d'une image comportant quatre pattes et une queue doit être étiqueté "chien". Bien entendu, le programme n'a pas connaissance des étiquettes "quatre pattes" ou "queue"; il va simplement chercher des motifs de pixels dans les données numériques. À chaque itération, le modèle prédictif créé par l'ordinateur devient plus complexe et plus précis. À la différence du bébé, qui prendra des semaines voire des mois pour comprendre le concept de “chien”, un programme informatique utilisant des algorithmes d’apprentissage profond peut trier des millions d’images en quelques minutes. Ce qui lui permet de détecter des chiens avec un taux d’erreur très faible.

Mais les premiers neural networks ne pouvaient simuler qu'un nombre très limité de neurones à la fois, de sorte qu'ils ne pouvaient pas reconnaître des motifs d'une grande complexité. La recherche a donc stagné dans les années 1970. Au milieu des années 1980, Geoffrey Everest Hinton et d'autres ont contribué à susciter un regain d'intérêt pour les neural networks avec des modèles dits "profonds", qui utilisaient de nombreuses couches de neurones artificiels. Mais la technique nécessitait encore une forte implication humaine: les programmeurs devaient étiqueter les données avant de les transmettre au réseau. Et la reconnaissance complexe de la parole ou de l’image nécessitait plus de puissance informatique.

Au cours de la dernière décennie, des avancées fondamentales ont vu le jour en grande partie grâce à la croissance exponentielle du “Big

Data", à la puissance de calcul des ordinateurs et aux améliorations apportées aux formules mathématiques.

Les informaticiens peuvent désormais modéliser beaucoup plus de couches de neurones virtuels. DeepMind de Google, sur lequel des dizaines de millions de vidéos YouTube ont été montrées, s'est avéré presque deux fois plus efficace que tous les efforts de reconnaissance d'image précédents, pour identifier des éléments tels que des chats. Google a également utilisé cette technologie pour réduire le taux d'erreur sur la reconnaissance vocale dans ses dernières versions Android sur smartphone.

En octobre 2012, le directeur de la recherche de Microsoft, Rick Rashid, a impressionné beaucoup de personne lors d'une conférence en Chine avec la démonstration d'un logiciel transcrivant ses mots en anglais avec un taux d'erreur de 7%, traduits en chinois puis simulant sa propre voix en mandarin. Ce même mois, une équipe d'étudiants a remporté un concours organisé par un grand laboratoire pour 63

identifier des molécules pouvant conduire à de nouveaux médicaments. Le groupe a utilisé le deep learning pour cibler les molécules les plus susceptibles de se lier à leurs cibles.

Le deep learning et les neural networks sont majoritairement utilisés dans les domaines tels que reconnaissance automatique de langage, traduction instantanée écrite et orale, reconnaissance d'image, restauration d'image, classification d'objets d'arts visuels, création de médicaments et modélisation toxicologique, système de recommandation, détection de fraude financière ou encore bio-informatique.

64

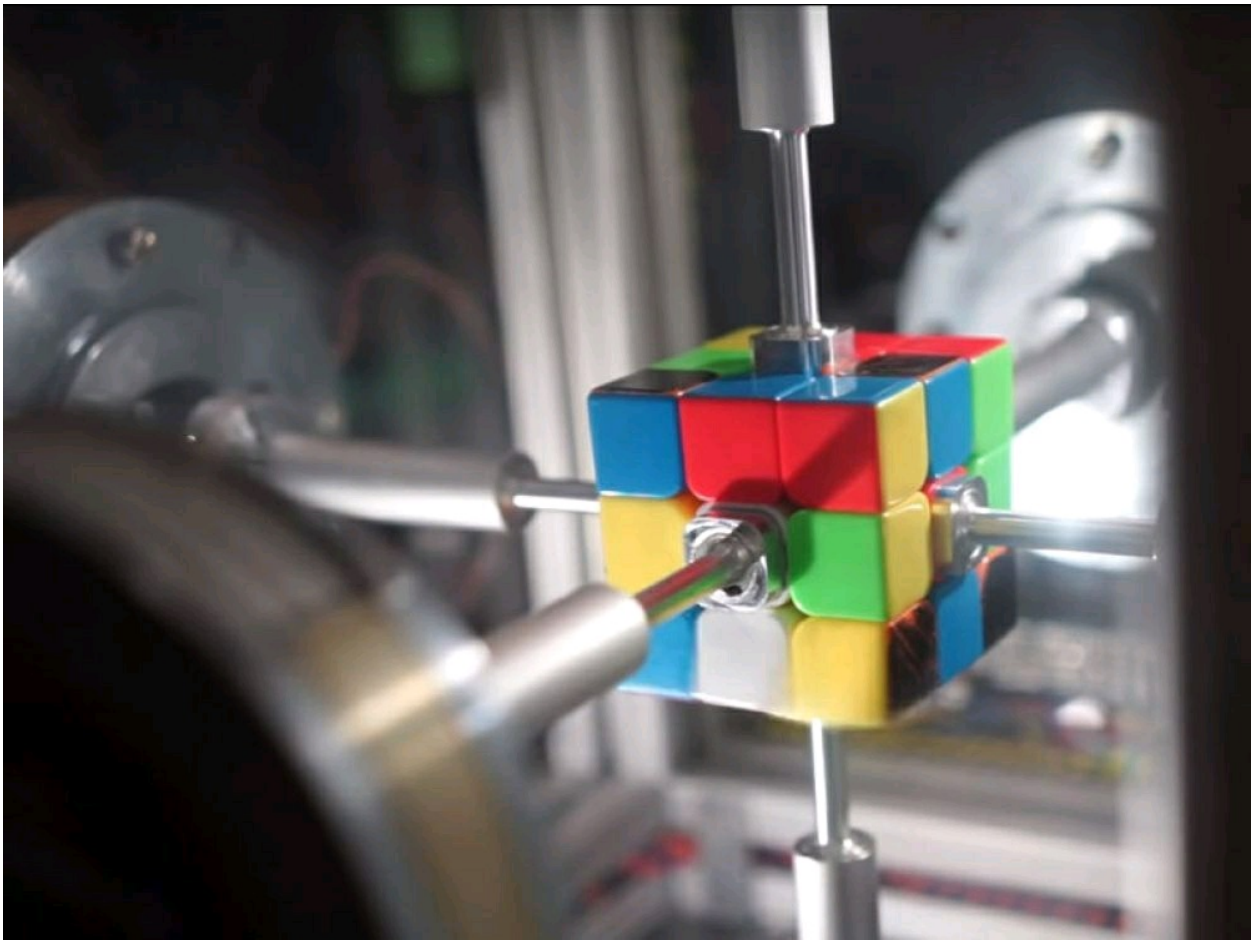
## 2. Aujourd'hui : Intelligence artificielle limitée

### Les progrès fulgurants

Comme nous l'avons vu, les améliorations exponentielles dans la mémoire et la puissance de calcul informatique se sont traduite en progrès spectaculaire dans le domaine de l'intelligence artificielle. Mais il a

fallu beaucoup de temps avant que le machine learning arrive à maturité. D'une manière générale, les intelligences artificielles ont longtemps été cantonnées dans une catégorie où elles étaient très efficaces lorsque les règles étaient précises et limitées. Prenons l'exemple du rubik's cube. Ce célèbre casse-tête a la réputation d'être relativement compliqué et le résoudre démontre une certaine preuve d'intelligence. Mais les règles sont précises, peu nombreuses et l'objectif est unique : avoir les six faces alignées par couleur. Depuis plusieurs années, de nombreuses équipes ont essayé de concevoir une machine capable de résoudre un rubik's cube et forcément, avec la bonne dose de puissance de calcul et une vitesse d'exécution extrêmement rapide, une machine peut surpasser un humain. L'humain le plus rapide arrive à résoudre le puzzle en 4,59 secondes. En novembre 2016, ce record a été battu. Le robot "Sub1 reloaded" l'a fait en 0,63

seconde. (Lien de la vidéo). En 2018, deux jeunes inventeurs ont fait même mieux en bricolant une machine capable de résoudre le Rubik's cube en 0,38 seconde. (Lien de la vidéo). Ce record ne pourra plus jamais être battu par un humain.



Cette machine peut résoudre un Rubik’s cube en 0,38 seconde. ©BEN KATZ CC-

BY-SA-4.0

Quand Deep Blue d’IBM est devenu champion du monde d'échecs en battant Garry Kasparov en 1997, ses principaux avantages résidaient dans la mémoire et le calcul, mais pas dans l’apprentissage. Son intelligence artificielle a été créée de toute pièce par une équipe d’êtres humains, et la raison principale pour laquelle Deep Blue pouvait dépasser ses créateurs était sa capacité à calculer plus rapidement et donc analyser plus de positions potentielles à chaque tour. Quand le superordinateur d’IBM, Watson, a détrôné le champion du monde dans

le jeu télévisé Jeopardy, ce n'était pas grâce à un système d'apprentissage. Mais bien une question de compétences programmées sur mesure, une mémoire supérieure et une vitesse d'exécution surhumaine. On peut dire la même chose de la plupart des avancées du 20e siècle en intelligence artificielle et robotiques.

Mais au 21e siècle, les percées se font bien plus grâce aux progrès dans le machine learning que grâce à une puissance de calcul élevée (même si ce dernier point joue toujours un rôle important).

66

**Résoudre un Rubik's cube sans aide initiale**

Un nouveau type de deep learning a appris comment résoudre un Rubik's cube sans aucune assistance humaine. Cette étape est importante et bien plus significative que les records du monde battu grâce à la "force brute" car la nouvelle approche aborde un problème important en informatique : comment résoudre des problèmes complexes lorsque l'aide initiale est minimale. Le programme reçoit les règles du jeu et joue ensuite contre lui-même. Le point crucial étant qu'il est récompensé en fonction de ses performances. Ce processus de récompense est extrêmement important, car il aide la machine à distinguer un bon coup d'un mauvais. En d'autres termes, cela aide la machine à apprendre. Résultat ? En 44 heures, le programme du nom de DeepCube a atteint le niveau d'un champion humain de la discipline. Assez impressionnant n'est ce pas ?

**Reconnaissance visuelle**

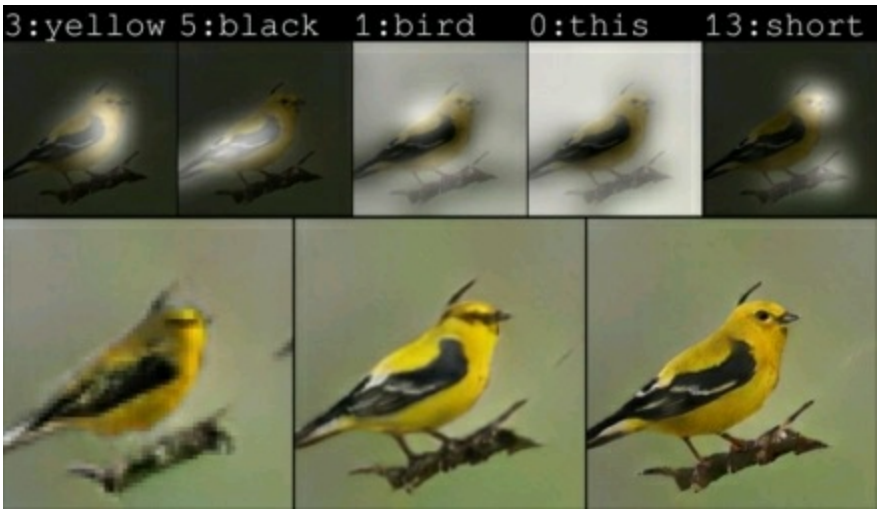
Afin de rendre une machine capable de comprendre le monde qui l'entoure, la technologie a été inspirée par la biologie. 80% des informations permettant aux humains de localiser leur espace et d'interagir avec leur univers passent par leurs yeux. Il a toujours été difficile pour un ordinateur de comprendre une image. C'est trivial pour un enfant de 4 ans, mais expliquer à un programme que les différents groupes de pixels colorés sur une image représentent, par exemple une montagne enneigée, est un problème qui a retenu l'attention des chercheurs depuis la création de la discipline de l'intelligence

artificielle. En d'autres termes, les ordinateurs n'ont jamais été capables de voir, de lire, ni d'entendre correctement. En 2004, le chercheur Jeff Hawkins déclara :" *Aucun ordinateur n'est capable de voir aussi bien qu'une souris*". Enfin, c'était vrai jusqu'à très récemment. Cela fait environ 5 ans que des percées majeures améliorent la reconnaissance visuelle des intelligences artificielles.

En 2011, une équipe de chercheurs de l'université de Stanford a construit le logiciel de reconnaissance d'images le plus efficace au monde à l'époque. Ce programme analysait un lot d'environ 50 000 images, et devait les classer dans 10

catégories, comme : "chiens", "chevaux" ou encore "camions". Il donna des réponses correctes environ 80% du temps. Sachant que des humains passant le test obtenaient un score moyen de 94%. À la suite de ce test, un étudiant de Stanford affirma qu'il serait difficile pour un ordinateur de faire mieux que 80% et qu'à la limite, il est envisageable de viser 85%, mais il faudra encore attendre longtemps avant d'obtenir un résultat égal à celui d'un humain. L'avenir ne lui donna pas raison puisqu'en 2014, une équipe de Google a conçu une intelligence artificielle capable d'écrire une courte description d'une image présentée. Le test appelé ImageNet est comparable à celui de l'université de Stanford, bien que plus compliqué. Des humains ont obtenu en moyenne 85%. Le programme de Google : 93,4%.

Describes without errors
Describes with minor errors
Somewhat related to the image
A person riding a motorcycle on a dirt road.
Two dogs play in the grass.
A skateboarder does a trick on a ramp.
A dog is jumping to catch a frisbee.
A group of young people playing a game of frisbee.
Two hockey players are fighting over the puck.
A little girl in a pink hat is blowing bubbles.
A refrigerator filled with lots of food and drinks.
A herd of elephants walking across a dry grass field.
A close up of a cat laying on a couch.
A red motorcycle parked on the side of the road.
A yellow school bus parked in a parking lot.




Un algorithme est capable d'étiqueter des images avec un haut taux de réussite.

Le constat est là : Les humains battaient facilement les logiciels de reconnaissance d'image en 2011. Trois ans plus tard, ce n'était plus le cas. Vous pouvez tester par vous même avec capionbot de Microsoft qui analyse le contenu d'une image que vous lui avez fourni (https://www.captionbot.ai/)

Microsoft a également mis au point un algorithme qui fait l'inverse en quelque sorte. C'est-à-dire qu'à partir de mots clés comme "oiseau", "Jaune", "Noir", l'intelligence artificielle génère une image.

Cet algorithme génère une image à partir de mots clés.

Microsoft appelle simplement cette nouvelle technologie “Drawing bot” pour le moment. Il peut générer des images d'animaux, des paysages, et même des choses bizarres comme des voitures volantes. Ce qui est le plus intéressante sur cette technologie c’est que les images générées ne sont peut-être même pas “Réel”.

68

L'oiseau créé dans l'image ci-dessus pourrait ne pas exister - ils ne sont qu'un rendu de ce qu’un oiseau jaune et noir est pour “l’imagination” de l’IA.

Donner des yeux à une intelligence artificielle permet des progrès dans de nombreux domaines. Les plus spectaculaires peuvent être observés dans la voiture autonome. Car ces technologies sont utilisées pour la détection des obstacles et la reconnaissance des panneaux, feux tricolores, voitures, piétons et autres. Les images proviennent d'une panoplie de caméras disposées autour de la voiture, et l'entraînement deep learning sur des millions d’heures de conduite a permis aux algorithmes une reconnaissance visuelle fiable.

Nous trouvons les algorithmes de reconnaissance d’image utilisés pour l'authentification d'un individu. L’iPhone X est un exemple notable grâce à la reconnaissance faciale en 3D. En matière de surveillance et de sécurité intérieure, la reconnaissance faciale est utilisée dans les contrôles aux frontières et dans la production de documents d’identité grâce à l’utilisation de caméras spécialisées. La reconnaissance d’Iris est également en net progrès avec une probable utilisation dans des applications mobiles.

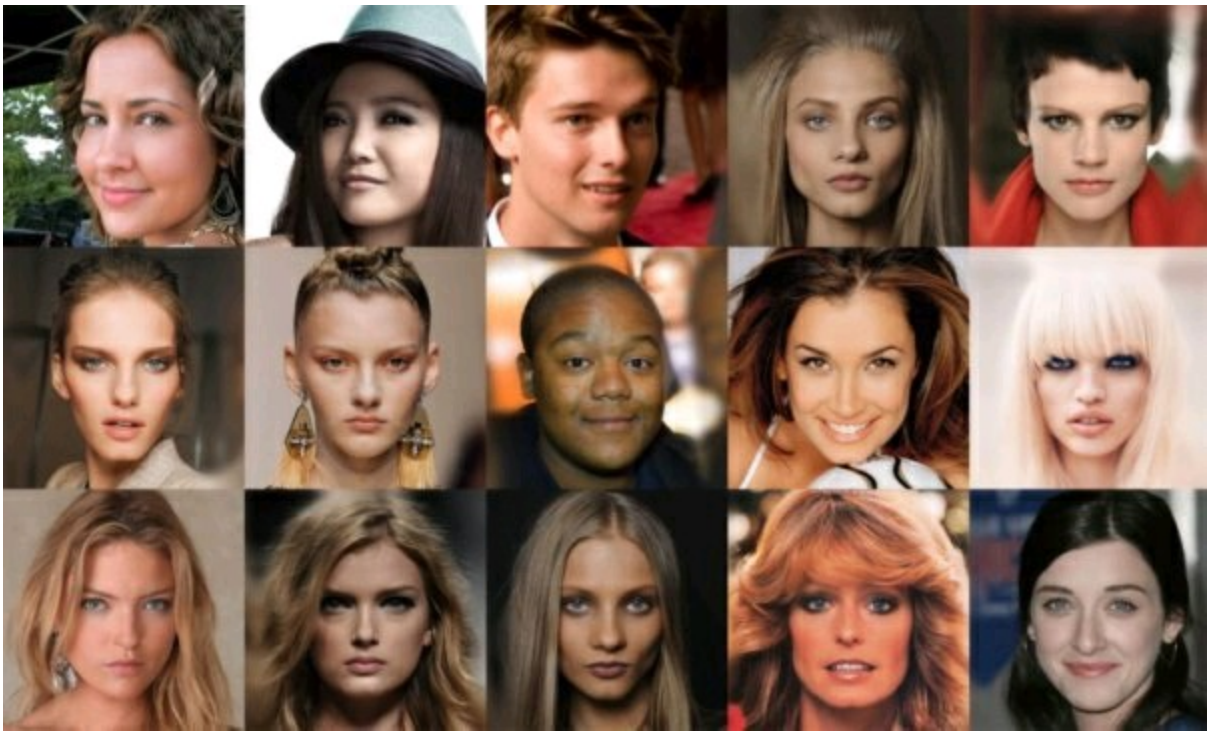
La reconnaissance visuelle donne à une intelligence artificielle la capacité de générer des créations artistiques. En nourrissant une IA avec des milliers d’oeuvre d’Art, le système est capable d’apprendre les styles artistiques des plus grands maîtres comme Picasso, Van Gogh ou Rembrandt. Ensuite, si on lui donne une photo quelconque, il est capable d’appliquer le style d’un artiste pour que la photo ait l’air d’avoir été peinte avec le style choisi. Cette technique s’appelle “Transfert de style”.

Ce qui suggère que d’une certaine façon, l’intelligence artificielle a “compris” le style artistique d’un peintre par intuition visuelle. Chose qui était complètement farfelue à imaginer il y a seulement vingt ans. La créativité et l’art sont des mots qui ne sont généralement pas souvent utilisés dans le domaine de l’intelligence artificielle. Du coup, ces programmes servent également aux musées et galeries d’art pour classer rapidement des oeuvres voire même identifier avec certitude si une peinture est un original ou une copie.

Si vous souhaitez vous amuser avec la technologie du transfert de style, vous aurez de bons exemples avec DeepDream de Google, sur deepdreamgenerator.com.

Au lieu de chercher le modèle idéal pour une séance photo, les photographes du futur pourront en générer un en utilisant l’intelligence artificielle. En 2017, NVIDIA, constructeur de processeur graphique, a publié un article décrivant des algorithmes capables de générer des "photos" artificielles qui semblent aussi réelles que la photo d’un humain.





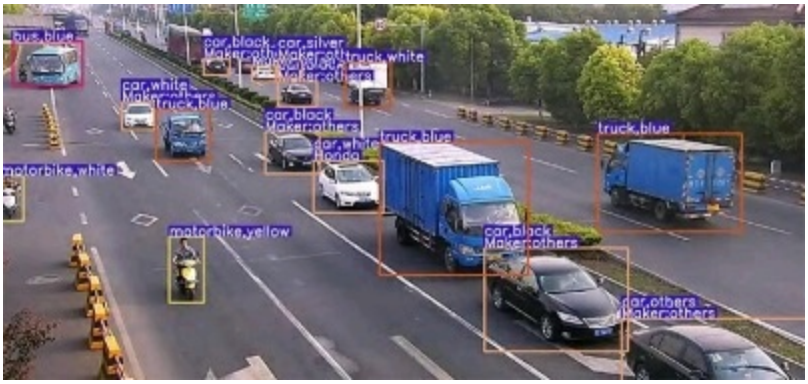
Pour son projet, NVIDIA a utilisé une base de données de visages célèbres.

Base de donnée de photos pour entraîner l'algorithme.

Après avoir été formé sur ces vrais visages, l'Intelligence artificielle était capable de commencer à produire des photos réalistes de personnes n’ayant jamais existés.

Photos de personnes entièrement générées par l'algorithme utilisant la technique GANs (generative adversarial networks).

70



Voir la video : https://tinyurl.com/y72wtute

La reconnaissance visuelle s'applique également au mouvement et à la vidéo.

Avec le deep learning, il est possible d'apprendre à des intelligences artificielles à reconnaître une personne sur une caméra de surveillance, ou de préciser si une vidéo concerne un match de football ou une course de voiture. Des algorithmes sont également capables d'identifier en temps réel des objets filmés, très utiles pour les voitures autonomes.

Les caméras de surveillance peuvent reconnaître de nombreux objets en

mouvement.

Lorsque les intelligences artificielles arrivent à mieux comprendre le monde réel, de nombreuses applications peuvent se répandre. Par exemple dans les maisons de retraite, en détectant si une personne est tombée dans sa chambre et nécessite de l'aide. Ou pour aider des personnes aveugles avec des outils décrivant ce qui se passe autour d'eux.

Des chercheurs travaillent également sur de la vidéo générative. L'idée est que, simplement en tapant une phrase, l'intelligence artificielle pourrait créer une vidéo de la scène décrite. Comme elle peut déjà le faire de façon sommaire avec une image.

Les chercheurs ont utilisé une variété d'activités faciles à définir pour leurs vidéos, principalement sportives comme "jouer au football", "courir dans la neige", "faire du surf". Le programme étudie ensuite ces vidéos et apprend à identifier chaque mouvement. Les chercheurs utilisent un processus en deux étapes pour créer la vidéo générative. La première consiste à générer la base de la vidéo à partir du texte saisi, c'est-à-dire une image qui donne la couleur d'arrière-plan et la disposition de l'objet de la vidéo souhaitée. Puis vient la deuxième étape, qui va juger le résultat. Pour une vidéo de "vélo dans la neige", la deuxième étape se penche sur la vidéo générée par rapport à une vidéo réelle de quelqu'un qui fait du vélo dans la neige et demande ensuite un travail de meilleure qualité. En utilisant des millions de connexions réseau, l'intelligence artificielle s'affine

constamment et finit par générer une vidéo passable qui représente le texte entré initialement. Le travail en est encore à ses premiers 71

stades, seulement capable de créer des vidéos d'une durée d'environ 1 seconde et qui sont de la taille d'un timbre-poste avec 64 par 64 pixels de résolution. Mais après plusieurs années de recherches, on pourra peut-être voir des films entièrement générés par une intelligence artificielle directement à partir d'un scénario préalablement écrit.

La fin du cinéma tel qu'on le connaît ? Ou plutôt un nouveau mouvement artistique ?

Il est aussi possible d'entraîner des intelligences artificielles à modifier une vidéo déjà existante. Nvidia, déjà auteur d'une percée sur la génération photoréaliste de personne, a démontré la possibilité de transformer une vidéo. Par exemple changer la météo d'un extrait de ensoleillé à pluvieux, et même enneigé. Ou passer une vidéo de jour à nuit.

https://tinyurl.com/ycbbs56e

https://tinyurl.com/y76ogptm

Le dernier exemple de ce que peut faire des intelligences artificielles dans le domaine de la reconnaissance visuelle provient de l'Université de Washington. En 2017, des chercheurs ont créé un nouvel outil qui prend des fichiers audio, les convertit en mouvements de bouche réalistes, puis les greffe sur une vidéo existante.

Le résultat final est une vidéo de quelqu'un qui dit quelque chose qu'il n'a pas dit au moment où la vidéo a été enregistrée.

https://tinyurl.com/y9r28o2n

Ce genre de transfert d'expression faciale est appelé "deepfake" et pose des questions troublantes sur les possibles manipulations d'une telle technologie. Il va devenir de plus en plus difficile de savoir ce qui est réel ou pas dans les médias. Voir la vidéo YouTube "Deepfake" par The Flares sur le sujet.

**Reconnaissance audio du langage**

La technologie de reconnaissance vocale est entrée dans la conscience collective assez récemment, avec le lancement des différentes intelligences artificielles des géants de la Silicon Valley (Google, Apple, Microsoft, Amazon). Nous sommes fascinés par l'idée que des machines peuvent nous comprendre. D'un point de vue anthropologique, nous avons développé la parole longtemps avant l'écriture et nous pouvons parler 150 mots par minute, comparé aux 40 mots dérisoires qu'une personne moyenne peut saisir en 60 secondes.

En fait, la communication vocale avec des appareils technologiques est en train de devenir si populaire et naturelle que nous pouvons nous demander pourquoi les entreprises les plus riches du monde ne nous apportent ces services que maintenant.

Jusque dans les années 1990, même les systèmes les plus performants étaient 72

basés sur la correspondance de modèles, où les ondes sonores étaient traduites en un ensemble de bits et stockées. Ces derniers étaient déclenchés lorsqu'un son identique était émis dans la machine. Bien sûr, cela signifiait qu'il fallait parler très clairement, lentement et dans un environnement sans bruit de fond pour avoir une bonne chance que le système reconnaisse les sons. Ce n'est qu'en 1997 que le premier "logiciel de reconnaissance vocale continue" (c'est-à-dire qu'il n'était plus nécessaire de faire une pause entre chaque mot) a été publié sous le nom de "Dragon NaturallySpeaking".

Capable de comprendre 100 mots par minute, il est toujours utilisé aujourd'hui, bien que sous une forme améliorée.

Mais le machine learning a fourni la majorité des avancées en matière de reconnaissance vocale au cours de ce siècle. Google a associé sa technologie la plus récente avec la puissance de l'informatique sur le cloud pour partager des données et améliorer la précision des algorithmes du machine learning. Cela a abouti au lancement de l'application Google Voice Search pour iPhone en 2008. Portée par d'énormes volumes de données d'apprentissage, l'application Voice Search a permis d'améliorer

considérablement les niveaux de précision des technologies de reconnaissance vocale. Google s'en est servi pour développer son algorithme Hummingbird, en parvenant à une compréhension beaucoup plus nuancée du langage utilisé. Ces éléments sont liés dans l'application "Google assistant", installée désormais sur près de 50% de tous les smartphones.

Ce fut Siri, l'entrée d'Apple sur le marché de la reconnaissance vocale, qui a d'abord séduit l'imagination du public. Grâce à des décennies de recherche, cet assistant numérique a apporté une touche d'humanité au monde stérile de la reconnaissance vocale. Enfin nous avions une voix qui n'était pas "robotique".

Microsoft a lancé Cortana, Amazon a lancé Alexa, et la bataille pour la suprématie de la reconnaissance vocale pouvait commencer.

Même si la route fut longue, les machines sont maintenant capables de comprendre la parole avec un taux de précision proche de 100%, et dans des centaines de langues. Les smartphones étaient à l'origine le seul lieu de résidence des assistants numériques tels que Siri et Cortana, mais le concept a été décentralisé ces dernières années.

À l'heure actuelle, l'accent est mis principalement sur des sortes de haut-parleurs activés par la voix, mais il s'agit essentiellement d'une stratégie de cheval de Troie. En prenant une place de choix dans la maison d'un consommateur, ces haut-parleurs sont la porte d'entrée vers la prolifération de dispositifs intelligents de l'Internet des objets.

Un Google Home ou Amazon Echo peut déjà être utilisé pour contrôler une vaste gamme d'appareils compatibles avec Internet, et beaucoup d'autres devraient rejoindre 73

la liste dans la décennie 2020. Ils incluront des réfrigérateurs intelligents, des fours, des miroirs, des détecteurs de fumée, etc.

À gauche : Google Home ©Daniel.Cardenas CC-BY-SA-4.0

À droite : Amazon Echo ©Asivechowdhury CC-BY-SA-4.0

Leur ambition est grandement aidée par le fait que la technologie est désormais utile dans l'accomplissement des tâches quotidiennes. Demandez à Alexa, Siri, Cortana ou Google quel temps il fera demain et il vous fournira un résumé détaillé.

C’est encore imparfait, mais la reconnaissance vocale a atteint un niveau de précision acceptable pour la plupart des utilisateurs, toutes les principales plates-formes affichant un taux d'erreur inférieur à 5%.

Maintenant que les utilisateurs surmontent la bizarrerie initiale de parler à leurs appareils, l'idée de dire à Alexa de faire bouillir la bouilloire ou de faire un expresso ne semble pas si étrange. Nous devrions donc nous attendre à voir des relations avec les assistants virtuels plus complexes à mesure que la technologie progressera. En 2018, Google a présenté une mise à jour de son Google assistant avec une démonstration impressionnante. L’intelligence artificielle reçoit pour tâche de réserve un rendez-vous chez le

coiffeur, et hop, elle s'exécute en appelant directement le salon de coiffure. La personne au bout du fil n'a pas conscience qu'elle parle à une IA. Il est tout à fait envisageable que les centres d'appel soient petit à petit remplacés par des intelligences artificielles. Si bien que lorsque l'on aura besoin d'un dépannage technique, ou de changer les formalités de son compte Amazon, nous ne saurons plus très bien si au bout de fil, nous parlons à un humain ou à une IA.

La voix humaine, avec toute sa subtilité et ses nuances, se révèle être une chose 74

extrêmement difficile à imiter pour une intelligence artificielle. Lorsque Siri, Alexa ou notre GPS communiquer avec nous, il est assez évident que c'est une machine qui parle. C'est parce que pratiquement tous les systèmes de synthèse vocale sur le marché reposent sur un ensemble préenregistré de mots, de phrases et d'énoncés (enregistrés à partir d'acteurs vocaux), qui sont ensuite liés pour produire des mots et des phrases complètes. Dans un effort pour donner un peu de vie aux voix automatisées, Lyrebird, une startup en IA, a développé un algorithme qui peut imiter la voix de toute personne, et lire un texte avec une émotion ou une intonation prédéfinie. Ce qui est incroyable, c'est qu'il peut le faire après avoir analysé quelques dizaines de secondes d'audio préenregistrées. Le résultat est loin d'être parfait, mais il est facile de reconnaître la personne dont la voix a été synthétisée.

Les chercheurs du géant de la technologie chinois Baidu, ont dévoilé leur dernière avancée de Deep Voice, un système développé pour le clonage de voix. Il y a un an, la technologie nécessitait environ 30 minutes d'audio pour créer un clip audio artificiel.

Maintenant, il peut créer de meilleurs résultats en quelques secondes seulement.

Inutile de dire que cette forme de synthèse vocale introduit une foule de problèmes éthiques et de problèmes de sécurité. Finalement, une version améliorée de ce système pourrait reproduire la voix d'une personne avec une précision incroyable, rendant pratiquement impossible à un humain de distinguer l'original de la copie. Des personnes sans scrupules

pourraient faire dire à un politicien des choses polémiques afin de changer le cours d'une élection, ou copier la voix d'une célébrité pour détruire sa carrière ou le faire chanter en menaçant de divulguer des extraits audio scandaleux.

Quels seront les recours possibles contre de tels actes ?

**Voitures autonomes**

Au cours des cinq dernières années, la conduite autonome est passée de "farfelu" à

"possible" à "inévitable" à "comment quelqu'un a-t-il pu penser que ce n'était pas possible?". Des expériences ont été menées sur l'automatisation des voitures depuis au moins les années 1920. Des essais prometteurs ont eu lieu dans les années 1950 et les travaux se sont poursuivis depuis. En 1969, John McCarthy (1927 - 2011), alias l'un des pères fondateurs de l'intelligence artificielle, décrit un véhicule similaire au véhicule autonome moderne dans un essai intitulé "Computer-Controlled Cars".

McCarthy fait référence à un "chauffeur automatique" capable de naviguer sur une voie publique via une caméra. Il écrit que les utilisateurs devraient pouvoir entrer une destination à l'aide d'un clavier, ce qui inciterait la voiture à les y conduire immédiatement. Aucun véhicule de ce type n'est construit, mais l'essai de McCarthy pose les bases pour les futures recherches.

Au début des années 1990, le chercheur Dean Pomerleau, rédige une thèse décrivant comment les neural networks pourraient permettre à un véhicule autonome 75

de prendre des images brutes de la route et de générer des commandes de pilotage en temps réel. Pomerleau n'est pas le seul chercheur travaillant sur les voitures autonomes, mais son utilisation des neural network s'avère beaucoup plus efficace que les alternatives.

En 2002, le gouvernement américain, via le DARPA, annonce une compétition offrant aux chercheurs un prix d’un million de dollars pour construire un véhicule autonome capable de parcourir 228 km à travers le désert de Mojave, à l’ouest des États-Unis. Lorsque le défi débute en 2004, aucun des 15 concurrents ne parvient à terminer le parcours. Cela porte un coup de froid dans l’objectif de construire de véritables voitures autonomes. Mais le défi est reconduit en 2005 avec cette fois, 5

véhicules qui termine la course. De bonnes augures pour la suite des recherches, mais pour monsieur et madame tout le monde, l'idée de se faire conduire par une série de 0

et 1 était ridicule pour ne pas dire terrifiante.

Une voiture autonome est développée pour le "Urban challenge" du DARPA de

l'édition 2007.

À partir de 2009, Google commence à développer son projet de voiture autonome appelé Waymo. Le projet fut initialement dirigé par Sebastian Thrun, ancien directeur du laboratoire d'intelligence artificielle de Stanford et co-inventeur de Google Street View. Mais la conduite autonome est une tâche beaucoup plus difficile qu'il n'y paraissait il y a quelques années, car les conducteurs humains ont beaucoup de 76

connaissance, pas seulement sur leur voiture, mais sur la façon dont les gens se comportent sur la route lorsqu'ils sont au volant. Pour parvenir à ce même type de compréhension, les voitures informatisées ont besoin de beaucoup de données. Et les deux sociétés qui ont le plus de données routières en ce moment sont Tesla et Waymo.

Tesla et Waymo tentent tous deux de collecter et de traiter suffisamment de données pour créer une voiture capable de conduire elle-même. Et ils abordent ces problèmes de manières très différentes. Tesla profite des centaines de milliers de leurs voitures sur la route en collectant des données réelles sur la performance de ces véhicules avec

la fonction Autopilot, son système semi-autonome actuel. Waymo de Google, utilise des simulations informatiques puissantes et alimente ce qu'il a appris de ceux-ci dans une flotte réelle plus petite d'environ 500 voitures.

Tesla (à gauche) et Waymo (à droite) sont deux acteurs majeurs du développement des voitures autonomes. ©Norsk Elbilforening CC BY 2.0 - ©Dllu CC BY-SA 4.0

Depuis octobre 2016, toutes les voitures Tesla sont équipées du matériel nécessaire pour permettre une conduite autonome complète à un niveau de sécurité 5, correspondant à l'autonomie complète. Le matériel comprend huit caméras et douze capteurs à ultrasons, en plus du radar orienté vers l'avant doté de capacités de traitement améliorées. Le système fonctionne en tâche de fond et renvoi des données à Tesla pour améliorer ses capacités jusqu'à ce que le logiciel soit prêt pour le déploiement via des mises à jour. L'autonomie totale n'est probable qu'après des centaines de millions de kilomètres d'essais et l'approbation des autorités.

Après quelques années, la technologie a atteint un point où aucun constructeur automobile ne pouvait l'ignorer. Depuis 2013, les principaux acteurs, notamment Audi, General Motors, Ford, Mercedes Benz, BMW, mais également Uber, utilisent 77

tous leurs propres technologies de voitures autonomes. Nissan s'engage sur une date de lancement en annonçant des voitures autonomes d'ici 2020.

Les voitures qui conduisent toutes seules sont maintenant partout. Elles parcourent les rues de Californie et du Michigan, de Paris et de Londres, de Singapour et de Beijing. Cette ruée vers l'or du XXIe siècle est motivée par l'appât du gain et l'instinct de survie. D'un côté, la technologie des voitures autonomes ajoutera 7 milliards de dollars à l'économie mondiale et sauvera des centaines de milliers de vies au cours des prochaines décennies. Mais en même temps, elle va dévaster l'industrie automobile, les chauffeurs de voitures, les chauffeurs de taxi et les routiers. Certaines personnes vont prospérer. La plupart en bénéficieront. Beaucoup vont perdre leurs emplois.

En décembre 2018, Waymo a lancé le premier service commercial de taxi autonome de l’histoire. Situé dans la ville de Phoenix aux États unis, c’est avant tout un test grandeur nature qui va servir de transition vers une adoption à grande échelle.

L’année 2019 devrait voir plusieurs offres de taxi autonome voir le jour.

**Alpha Go**

Le jeu de Go est né en Chine il y a 3 000 ans. Les règles du jeu sont simples: les joueurs placent à tour de rôle des pierres noires ou blanches sur un tableau, essayant de capturer les pierres de l'adversaire et créer des territoires. Go se joue principalement par l'intuition et le ressenti, et grâce à sa subtilité et sa profondeur intellectuelle, il a capté l'imagination humaine pendant des siècles. Il est considéré comme l’un des quatre arts chinois avec la calligraphie, la peinture chinoise et le guqin (instrument à cordes). Aussi simples que soient les règles, Go est un jeu d'une grande complexité. Il y a 10170 configurations de jeu possibles, plus que le nombre d'atomes dans l'univers observable, faisant de Go un jeu un googol fois plus complexe que les Échecs.

Le jeu de Go se joue sur une grille avec des billes blanches et noires.

Donc, concevoir une intelligence artificielle capable de battre un humain a Go était un défi bien plus difficile que pour les échecs. Pas étonnant qu'il ait fallu deux décennies de plus. La complexité de Go signifie qu'il a longtemps été considéré comme le plus difficile des jeux classiques pour l'intelligence artificielle. Malgré des décennies de travail, les programmes informatiques les plus performants ne pouvaient jouer qu'au niveau des amateurs humains. Certains experts ont longtemps pensé que c'était une tâche impossible à réaliser. Pourtant, Google a réussi avec le programme DeepMind en utilisant son intelligence artificielle appelée Alpha Go qui a battu en 2016 le champion coréen Lee Sedol, considéré comme l'un des meilleurs joueurs de l'histoire. Le match s'est déroulé en 5 parties en mars 2016 dans un Hôtel à Séoul.

Voici les témoignages de ce que pensait Lee Sedol à propos d'AlphaGo avant, puis après les matchs contre l'intelligence artificielle :

- Octobre 2015 : “*Basé sur ce que j’ai vu de son niveau… je pense que je vais gagner de manière écrasante. ”*

- Février 2016 : ‘ *J'ai entendu dire que l’IA de Google DeepMind devenait de plus en plus forte, mais je suis convaincu que je peux gagner au moins sur cette partie.* "

- 9 mars 2016 : “*J'ai été très surpris parce que je ne pensais pas perdre. ”*

- 10 mars 2016 : “*Je suis sans voix… Je suis sous le choc. J’admets que… le* 79

*troisième match ne sera pas facile pour moi. ”*

- 12 mars 2016 : “*Je me suis senti impuissant*”.

En 6 mois, Lee Sedol a vu sa confiance passer d’imbattable à impuissante face à AlphaGo. Moins d’un an après avoir battu Lee Sedol, une version améliorée d’AlphaGo a joué contre les vingt meilleurs joueurs du monde sans perdre un seul match.

L'algorithme d'AlphaGo combine des techniques de machine learning et de recherche arborescente, associées à une formation poussée, à la fois sur des jeux humains et informatiques. Il a d'abord été formé pour imiter le jeu humain en tentant de faire correspondre les mouvements de joueurs experts avec des jeux historiques enregistrés, en utilisant une base de données d'environ 30 millions de mouvements sur 160 000 parties. Une fois qu'il a atteint un certain degré de compétence, il a été entraîné à jouer un grand nombre de parties contre des instances de lui-même. Ce mariage entre intuition et logique a donné naissance à des mouvements très créatifs, défiant des millénaires d'expérience humaine.

Google DeepMind a ensuite fait évoluer AlphaGo en AlphaGo Zero. Les versions précédentes d'AlphaGo s'entraînaient initialement sur des milliers de parties amateurs et professionnels pour apprendre les différentes stratégies. AlphaGo Zero a ignoré cette étape et a simplement appris à jouer contre lui-même. De manière surprenante, il a rapidement dépassé le niveau humain et a vaincu la version d'AlphaGo par 100

parties à 0. En trois jours il a atteint le niveau de la version qui a battu Lee Sedol. En 21 jours, il a atteint le niveau de AlphaGo master qui a battu 60 joueurs professionnels de niveau international. En 40 jours, il dépassa toutes les versions précédentes d'AlphaGo et devint le meilleur joueur de Go de l'histoire. Tout ça sans aucun apprentissage en amont des règles du jeu, sans intervention humaine, sans analyser des anciennes parties.

Cela est possible en utilisant une nouvelle forme d'apprentissage par renforcement, dans laquelle AlphaGo Zero devient son propre professeur. Le système commence avec un réseau neuronal qui ne sait rien du jeu de Go. Il joue ensuite contre lui-même, en combinant ce réseau neuronal avec un puissant algorithme de recherche. Pendant qu'il joue, le réseau neuronal est réglé et mis à jour pour prédire les mouvements, ainsi que l'éventuel gagnant des parties.

Ce réseau neuronal mis à jour est ensuite recombiné avec l'algorithme de recherche pour créer une nouvelle version plus puissante d'AlphaGo Zero, et le processus recommence. À chaque itération, les performances du système s'améliorent légèrement et la qualité des parties augmente, ce qui permet d'obtenir des réseaux 80

neuronaux de plus en plus précis et des versions toujours plus puissantes d'AlphaGo Zero. Cette technique est plus puissante que les versions précédentes d'AlphaGo, car elle n'est plus limitée par les limites de la connaissance humaine. Au lieu de cela, l'intelligence artificielle est capable d'apprendre à partir du joueur le plus fort au monde: AlphaGo lui-même. Cette technique fut ensuite utilisée pour apprendre les échecs à AlphaZero qui a atteint des résultats impressionnants en quelques heures, cherchant mille fois moins de positions que les autres programmes, sans aucune connaissance du jeu si ce n'est les règles. Demis Hassabis de DeepMind, un joueur d'échecs lui-même, a surnommé le style de jeu d'AlphaZero "d'extra-terrestre".

**Jeux vidéo**

En 2015, Google DeepMind a publié un système d'IA utilisant du deep learning capable de maîtriser des dizaines d'anciens jeux vidéos

ATARI, sans aucune instruction initiale. Au départ, l'intelligence artificielle commence à jouer et se rend compte qu'en fonction de certaines actions, le score augmente, ce qui est jugé positif.

Donc elle apprend petit à petit à maximiser les façons de faire augmenter le score.

Après quelques heures d'entraînement, elle atteint un niveau imbattable par un humain.

Par exemple, sur le jeu "Breakout", où le but est de déplacer une planche en bas de l'écran pour faire rebondir une balle qui fait percuter des briques de couleurs en haut de l'écran. Chaque brique détruite donne des points. Au départ, l'Intelligence artificielle était vraiment mauvaise. Elle déplaçait la planche au hasard de droite à gauche sans se soucier de la balle. Après un certain temps, elle semblait comprendre que déplacer la palette vers la balle pour la faire rebondir était une bonne idée, même si elle manquait toujours la plupart du temps. Mais elle a continué à s'améliorer et après 2 heures de pratique, elle était capable de toucher la balle à tous les coups, quelle que soit la rapidité avec laquelle elle s'approchait. Mais le véritable exploit, c'est lorsque l'intelligence artificielle a compris une incroyable stratégie pour maximiser le score en visant toujours le coin supérieur gauche pour percer un trou dans le mur de brique et laisser la balle se coincer derrière afin de rebondir toute seule et engranger des points. On peut clairement dire que c'est une technique vraiment intelligente à faire. En effet, Demis Hassabis, à la tête de DeepMind, a déclaré que l'équipe de programmeur ne connaissait pas cette astuce. L'intelligence artificielle l'a apprise toute seule : https://tinyurl.com/y76rrcb7

Le deep learning a donné d'excellents résultats dans des jeux solos de plus en plus complexes et dans des jeux tour par tour à deux joueurs comme les échecs ou Alpha Go. Cependant, le monde réel contient plusieurs "joueurs", chacun apprenant et agissant indépendamment pour coopérer et rivaliser avec d'autres, et les environnements reflétant ce degré de complexité restent un défi ouvert. Mais en 2018, 81

une intelligence artificielle a atteint pour la première fois le niveau humain dans un jeu vidéo multijoueur, à savoir Quake III Arena dans mode capture de drapeau.

Pendant le jeu, l'intelligence artificielle démontre des comportements de type humain, tels que la navigation, le suivi et la défense, basés sur une très grande connaissance du jeu. Lors d'un tournoi, les agents artificiels ont dépassé le taux de victoires des meilleurs joueurs humains en tant que coéquipiers et adversaires. Ces résultats démontrent une augmentation significative des capacités des agents artificiels, nous rapprochant de l'objectif de répliquer l'intelligence humaine.

Autre exemple avec le célèbre jeu multijoueur DOTA 2. Les équipes de l'organisation Open AI, fondée par Elon Musk, travaille sur la mise au point d'une intelligence artificielle capable de battre les champions de DOTA 2. Chaque jour, le programme joue contre lui-même l'équivalent de 180 années de partie non-stop. Lors d'un tournoi en août 2018, l'intelligence artificielle a battu plusieurs champions avant de finalement s'incliner. Mais personne n'est dupe face à cette victoire

humaine. Le jour où une IA sera imbattable à DOTA 2 est très proche.

Compétition internationale de DOTA 2 ©Jakob Wells CC-BY-2.0



## 2. Aujourd'hui : Intelligence artificielle limitée

### Un monde sous les IA limitées

Après avoir passé en revue l'histoire et l'état actuel de l'intelligence artificielle, et vu que vous devez avoir une meilleure compréhension du sujet, à présent, on peut se pencher sur l'avenir. Activité que j'apprécie tout particulièrement avec The Flares.

Tout d'abord, nous allons tenter d'anticiper les bouleversements possibles que l'intelligence artificielle limitée pourrait engendrer en fonction de ses éventuels progrès et évolutions dans un futur proche. Car le secteur de l'intelligence artificielle représente l'un des marchés économiques et d'investissement le plus en croissance.

Selon le dernier rapport d'étude de marché par l'organisme de recherche

"MarketsandMarkets", le marché de l'intelligence artificielle devrait passer de 21,46

milliards USD en 2018 à 190,61 milliards USD d'ici 2025, soit un taux de croissance annuel moyen de 36,62% entre 2018 et 2025.

L'intelligence artificielle est considérée comme le prochain grand développement technologique, comparable à la révolution industrielle, l'ère de l'informatique et l'émergence des smartphones, qui ont tous façonné le monde d'aujourd'hui. La région nord-américaine devrait dominer l'industrie en raison de financements publics élevés, de la présence d'acteurs de premier plan et d'une base technique solide. Mais l'Asie, avec en son coeur la Chine, sera un concurrent très important qui pourrait précipiter la recherche en une sorte de course à l'armement. Les progrès en

reconnaissance d'image et du langage entraînent une croissance du marché de l'intelligence artificielle, car une technologie de reconnaissance d'image améliorée est essentielle pour les drones, les voitures autonomes et la robotique. Les deux principaux facteurs favorisant la croissance du marché sont les technologies émergentes comme le machine learning et le neural network ainsi que la croissance du big data et de la puissance de calcul.

**Automatisation et emploi**

Le chômage technologique est la perte d'emplois provoquée par un changement technologique. Un tel changement comprend généralement l'introduction de machines ou des processus automatisés plus intelligents qui permettent d'économiser de la main-d'œuvre. De même que l'automobile a progressivement rendu obsolètes les chevaux, les emplois humains ont également été affectés au cours de l'histoire moderne. Les exemples historiques incluent les artisans tisserands réduits à la 83

pauvreté après l'introduction de métiers à tisser mécanisés. Un exemple contemporain de chômages technologiques est le déplacement des caissiers de détail par les caisses libre-service.

Peut-être le premier exemple d'une personne discutant du phénomène du chômage technologique remonte à Aristote (384 av. J.-C - 322 av. J.-C.) il y a 2400 ans. Il a spéculé dans le premier tome de “Politique” que si les machines pouvaient devenir suffisamment avancées, le travail humain ne serait plus nécessaire. On peut dire que le monsieur a eu du flair sur l’avenir ! Depuis la publication de leur livre en 2011 “Race Against The Machine”, les professeurs du MIT, Andrew McAfee et Erik Brynjolfsson, ont été au premier plan parmi les personnes préoccupées par le chômage technologique. Les deux professeurs restent cependant relativement optimistes, déclarant que *"la clé pour gagner la course n'est pas de rivaliser contre les machines, mais de rivaliser avec des machines*".

Une usine automatisée par des robots - ©ICAPlants CC BY-SA 3.0

Un nouveau rapport prédit qu’en 2030, 800 millions d’emplois dans le monde pourraient être menacés à cause de l’automatisation. L'étude, compilée par le McKinsey Global Institute, indique que les progrès de l'IA et de la robotique auront un effet radical sur la vie professionnelle, comparable à l'abandon des sociétés 84

agricoles pendant la révolution industrielle. Aux États-Unis, entre 39 et 73 millions d'emplois devraient être automatisés, soit environ un tiers de l'effectif total.

Mais le rapport indique également que, comme par le passé, la technologie ne sera pas une force purement destructive. De nouveaux emplois seront créés, les rôles existants seront redéfinis, et les travailleurs auront la possibilité de changer de carrière. Le défi particulier de cette génération, affirment les auteurs, consiste à gérer la transition. L'inégalité des revenus est susceptible de s'accroître, ce qui pourrait entraîner une instabilité politique et les personnes qui doivent se recycler pour de nouvelles carrières ne seront pas des jeunes, mais des professionnels d’âge moyen.

Si on regarde uniquement ce qu’une technologie comme les voitures

autonomes va engendrer comme conséquence sur le monde du travail, c'est potentiellement des millions d'emplois qui peuvent disparaître. Chauffeur de bus, taxi, routiers… toutes les personnes qui gagnent leur vie en conduisant une voiture ou un camion pour faire simple. Une étude du département du commerce et de l'économie des États unis révèle qu'en 2015, 15,5 millions de travailleurs américains occupaient des emplois qui pourraient être affectés (à des degrés divers) par l'introduction de véhicules automatisés. Cela représente environ un travailleur sur neuf. Bien que la technologie se soit développée rapidement ces dernières années, le consensus est que l'adoption généralisée de véhicules "entièrement" automatisés prendra probablement une bonne décennie supplémentaire. Dès lors, les entreprises commerciales utilisant actuellement des véhicules automobiles et des conducteurs humains pourront adopter de plus en plus les véhicules autonomes dans le cadre de leur processus de production. L'étude a utilisé les données disponibles sur les activités des différentes professions pour déterminer celles qui sont le plus liées à la conduite et qui, par conséquent, sont plus susceptibles d'être affectées par l'utilisation future des véhicules autonomes par les entreprises. Cette analyse montre que l'adoption des véhicules autonomes peut avoir un impact considérable sur une partie des emplois dans l'économie. La mesure dans laquelle ils pourraient éliminer certaines professions, entraînant une perte d'emploi, tout en modifiant la combinaison des tâches impliquées dans d'autres professions, n'est toujours pas claire.

Néanmoins, les résultats suggèrent que les travailleurs de certaines professions liés à la conduite pourraient avoir des difficultés à trouver un autre emploi. De manière générale, ils sont plus âgés, moins scolarisés et, dans la plupart des cas, ont moins de compétences transférables que les autres travailleurs, en particulier sur les compétences requises pour des tâches cognitives non répétitives.

Les voitures autonomes vont probablement éliminer de nombreux emplois -

©David Castor CC0

Les changements ne toucheront pas tout le monde également. Seulement 5% des

emplois actuels seront complètement automatisés si la technologie de pointe d'aujourd'hui est massivement adoptée, alors que dans 60% des emplois, un tiers des activités sera automatisé. Citant une commission du gouvernement américain des années 1960 sur le même sujet, les chercheurs de McKinsey résument: "La technologie détruit les emplois, mais pas le travail". Elle examine, par exemple, l'effet de l'ordinateur avec la création de 18,5 millions de nouveaux emplois, même en tenant compte des emplois perdus. (La même chose pourrait ne pas être vraie pour les robots industriels, qui suggèrent selon les rapports précédents la destruction des emplois en général.)

Les prévisions économiques ne sont pas une science exacte et les

chercheurs de McKinsey tiennent à souligner que leurs prédictions ne sont que cela. Le chiffre de 800 millions d'emplois perdus dans le monde, par exemple, n'est que le scénario le plus extrême possible, et le rapport suggère également une estimation moyenne de 400 millions d'emplois.

Néanmoins, cette étude est l'une des plus complètes de ces dernières années.

Prenant en compte l'évolution de la modélisation dans plus de 800 professions et dans 46 pays, représentant 90% du PIB mondial. Six pays sont également analysés en 86

détail - les États-Unis, la Chine, l'Allemagne, le Japon, l'Inde et le Mexique - ces pays représentant un éventail de situations économiques différentes.

Le rapport souligne que les effets de l'automatisation sur le travail seront différents d'un pays à l'autre. Les économies développées comme les États-Unis et l'Allemagne seront probablement les plus touchées par les changements à venir, car des salaires moyens plus élevés favorisent l'automatisation. En Amérique, le rapport prédit que l'emploi dans les industries comme les soins de santé augmentera à mesure que la société affrontera une population vieillissante. Alors que les emplois spécialisés qui impliquent un travail physique ou le traitement de données sont les plus exposés à l'automatisation.

Dans les économies développées telles que les États-Unis, l'automatisation risque également d'accroître les inégalités. Les emplois créatifs et cognitifs bien rémunérés seront privilégiés, tandis que la demande de professions moyennement qualifiées et peu qualifiées diminuera. Selon McKinsey, le résultat sera un "marché du travail à deux niveaux". Les rapports précédents sont arrivés à la même conclusion, constatant que les individus ayant des revenus plus élevés sont plus à même de s'adapter à un marché du travail en mutation et que la mobilité sociale en souffrira.

Toutefois, les pires effets de cette transition peuvent être atténués si les gouvernements jouent un rôle actif. Michael Chui, l'auteur principal du

rapport, a comparé le niveau d'action nécessaire au plan Marshall - une initiative américaine qui a injecté quelque 140 milliards de dollars en Europe occidentale après la Seconde Guerre mondiale, aidant les pays à se reconstruire et à s'industrialiser.

Le rapport utilise la transition des États-Unis en dehors de l'agriculture comme exemple historique, soulignant que la diminution des emplois agricoles aux États-Unis s'accompagnait de dépenses importantes dans l'enseignement secondaire et de nouvelles lois imposant l'école obligatoire. En 1910, seuls 18% des enfants âgés de 14 à 17 ans allaient au lycée; en 1940, ce chiffre était de 73%. L'augmentation des travailleurs instruits qui en a résulté a contribué à créer une industrie manufacturière en plein essor et une classe moyenne florissante. Un effort similaire est nécessaire aujourd'hui, dit McKinsey, mais au cours des dernières décennies, les dépenses en formation et en soutien ouvrier ont diminué. La conclusion du rapport semble être la suivante: l'automatisation ne doit pas nécessairement être une catastrophe, mais seulement si la politique arrive à s'adapter et encourager une transition en douceur.

Comment préparer nos enfants pour le monde de demain ? Que faut-il enseigner aux enfants à l'école ? Il y a 50-100 ans, c'était plus simple, car entre deux ou trois générations, le monde n'évoluait que très peu. Un enfant allait à l'école et il apprenait 87

des techniques que ses parents avaient également apprises. Il n'y avait que très peu de nouveaux métiers. Donc il pouvait choisir une branche et il était sûr qu'il trouverait un emploi. On ne peut plus se baser sur le système précédent. Si on prend le cas d'un étudiant en informatique en 2010. Lorsque ses parents étaient à l'école, aucun professeur n'enseignait la programmation informatique, car il n'y avait aucun métier dans ce secteur. En l'espace de 25-30 ans, le progrès a engendré une pléthore de nouveau job. Et ce rythme du progrès augmente.

La vérité, c'est que personne ne sait ce qui se passera dans les années 2030-2040

et au-delà. Et encore moins à quoi ressemblera le marché du travail.

C’est bien la première fois dans l’histoire que cela arrive ! On ne peut plus simplement former des jeunes pour qu’ils aillent occuper certains métiers spécifiques dans les dix prochaines années. Car tout ce qu’ils vont apprendre sera peut-être à jeter par la fenêtre. Nous avons une certaine vision du monde qui consiste à séparer notre vie professionnelle en deux phases. La phase d’apprentissage où l’on étudie et développe un ensemble de compétence. Et une phase de travail où l’on applique ces compétences dans une discipline, un corps de métier jusqu’à la retraite. Aujourd’hui cette vision est dépassée

! Comment préparer les jeunes à un monde où de plus en plus d’emplois seront remplacés par des systèmes automatisés ? La mobilité sera essentielle. Mais elle sera également bien plus compliquée. Lors de la première révolution industrielle, lorsque des paysans n'étaient plus désirés, car des machines étaient capables de récolter bien plus de céréales, ils pouvaient facilement se reconvertir et aller travailler dans des usines. Après une légère formation, ils étaient capables d'évoluer dans une ligne d’assemblage. Ce sont deux métiers qui requièrent peu d’apprentissages et compétences. Des métiers “low skills” basés sur des tâches répétitives. Mais demain, lorsque des caissiers se feront licencier, car le supermarché passera en complète automatisation, ce sera bien plus compliqué pour eux de se reconvertir en développeur de monde virtuel. Lorsqu’un chauffeur de taxi de 45 ans sera remplacé par une voiture autonome, ce ne sera pas évident de devenir en quelques mois un spécialiste en machine learning. Passer d’un métier “low skill” à un autre métier “low skill” est un exercice bien plus simple que de passer de low skill à high skill, c’est-à-dire à compétences élevées. Et même ces métiers-là sont menacés par l’intelligence artificielle.

Les effets de l’innovation sur le marché du travail ont généralement nui principalement aux travailleurs peu qualifiés, tout en bénéficiant souvent à des travailleurs qualifiés. Selon des chercheurs tels que Lawrence F. Katz, c’était peut-être le cas pendant la majeure partie du vingtième siècle, mais au 19e siècle, les innovations sur le lieu de travail ont largement déplacé les artisans qualifiés par exemple. Alors que l’innovation du 21e siècle a remplacé certains emplois non qualifiés, d’autres professions peu qualifiées restent résistantes à l’automatisation,

tandis que des professions qualifiées comme assistants légaux (clercs) sont de plus en plus remplacées par des programmes informatiques autonomes.

Certaines études récentes, comme un article publié en 2015 par Georg Graetz et Guy Michaels, ont montré que l'innovation dans les robots industriels stimule la rémunération des travailleurs hautement qualifiés tout en ayant un impact plus négatif sur les travailleurs de compétences faibles à moyennes. Un rapport publié en 2015 par Carl Benedikt Frey, Michael Osborne et Citi Research, a reconnu que l'innovation avait surtout perturbé les emplois de niveau intermédiaire, tout en prévoyant que l'impact de l'automatisation dans les dix prochaines années impactera principalement les compétences peu qualifiées. En revanche, d'autres voient même les travailleurs humains qualifiés obsolètes. Carl Benedikt Frey et Michael A Osborne de l'université d'Oxford ont prédit que l'automatisation pourrait faire perdre la moitié des emplois sur les 702 professions évaluées par leur recherche. Ils ont constaté une forte corrélation entre les études, les revenus et la capacité à être automatisé. Les emplois de bureau et les services étant parmi les plus menacés. En 2012, le cofondateur de Sun Microsystems, Vinod Khosla, a prédit que 80% des emplois de docteurs seraient perdus au cours des deux prochaines décennies grâce à un logiciel de diagnostic médical automatisé disponible en masse.

Une autre approche afin d'anticiper les compétences sur lesquelles les humains apporteront une valeur plus importante que les IA est de trouver des activités où nous voulons que les humains restent responsables des décisions importantes. Comme les juges, les PDG, et les dirigeants gouvernementaux. Ou bien apportent des relations interpersonnelles profondes. Car même si ces tâches peuvent être automatisées, il y a de fortes chances pour que la valeur de l'interaction humaine prédomine.

Que faut-il enseigner aux enfants ? On peut commencer par des notions comme une forte ténacité à accepter le changement, faire preuve d'adaptabilité, de résilience et surtout, de toujours apprendre. L'éducation ne s'arrête pas après l'université. Ce temps-là est terminé et de toute façon, c'est une idée qui ne fait aucun sens. On apprend jusqu'à la fin

de notre vie. Une des qualités qu'il faudra posséder dans ce monde en constant changement, c'est la capacité de se réinventer soi-même. Il ne faudra pas avoir peur de sauter sur des nouveaux domaines d'apprentissage, changer de branche. Et ce n'est pas simple. En particulier lorsqu'on prend de l'âge, car prendre des risques signifie sortir de sa zone de confort.

Donc pour résumer, il y a deux types de tâches qu'un être humain peut faire. Des tâches physiques, et des tâches cognitives. L'innovation technologie dans le monde du travail a toujours été cantonné dans le camp des tâches physiques. La roue, la charrue, le moulin, le métier à tisser, la poulie, les machines à vapeur, les robots industriels, 89

etc. Et même si cela a supprimé des emplois humains, une mobilité est possible vers un autre secteur possédant les mêmes types de compétences.

Un engin à vapeur fut l'une des premières technologies de la révolution

industrielle au 18e siècle, qui a supprimé des emplois physiques - CC BY-SA 3.0

Cette fois-ci, l'innovation technologique est capable d'automatiser bien plus d'emploi si bien qu'il devient de plus en plus difficile d'avoir une mobilité horizontale. Et une mobilité verticale, de peu qualifié vers très qualifié, est généralement limité à un petit nombre d'individus. Ce qui signifie un chômage qui explose. Pour la première fois également, l'automatisation s'attaque aux tâches cognitives qui requièrent historiquement un haut niveau d'étude.

Voici une liste non exhaustive de profession qui pourrait voir des systèmes automatisés remplacer des emplois humains à court ou moyen terme :

**- Ressources humaines :** Des algorithmes sont développés pour automatiser les décisions de gestion, y compris l'embauche et le licenciement des employés.

**- Centres d'appel :** La reconnaissance vocale et l'amélioration des voix artificielles sont capables de prendre en charge de nombreuses requêtes de clients.



**- Avocats :** En s'entraînant sur des millions de documents et dossiers juridiques existants, un algorithme d'apprentissage automatique peut apprendre à identifier les sources appropriées dont un avocat a besoin pour élaborer un dossier, souvent avec plus de succès que les humains. Par exemple, JPMorgan utilise un logiciel appelé Contract Intelligence, ou COIN, qui peut en quelques secondes effectuer des tâches d'examen des documents qui ont nécessité 360 000 heures d'aide juridique.

**- Journalistes :** Les bots créés par des sociétés telles que Narrative Science et Automated Insights proposent déjà des articles sur le business et le sport à des clients tels que Forbes et Associated Press. Dans une interview accordée au Guardian en juin 2015, le cofondateur de Narrative Science, Kris Hammond, a prédit que 90% du journalisme sera informatisé d'ici 2030, et qu'un bot gagnera le prix Pulitzer.

**- Thérapeutes :** Des "robots sociaux" ressemblant à des êtres humains sont

déjà utilisés avec des enfants autistes. Les animaux de compagnie thérapeutiques fournissent de la compagnie aux personnes âgées atteintes de démence. Tandis que l'armée américaine utilise un thérapeute virtuel pour dépister les patients atteints de syndromes post-traumatiques.

- **Enseignants :** Des logiciels tels que Aplia permettent aux professeurs d'université de gérer leurs cours pour des centaines d'étudiants à la fois et des cours en ligne regroupent des milliers de disciplines. Et des robots physiques réels sont utilisés pour enseigner l'anglais aux étudiants au Japon et en Corée.

- **Auteurs :** En janvier 2015, Watson d'IBM, l'intelligence artificielle qui a gagné le jeu TV Jeopardy, a publié un livre contenant 65 recettes de cuisine élaborées. Ce qui ouvre la porte à de nombreux autres livres du même style qui pourraient être écrits par des IA.

- **Livreurs :** Aloft Hotels expérimente un robot-hôtelier appelé "Botlr" pour livrer des articles de toilette dans votre chambre. Le robot de livraison de Starship Technologies, qui ressemble à un aspirateur robotique, peut transporter des aliments et des colis vers des destinations situées dans un rayon proche. Et en décembre 2017, Amazon a livré son premier colis à un client utilisant un drone.

Amazon Prime Air s'engage à livrer des colis pesant moins de trois kilogrammes en 30 minutes ou moins le plus tôt possible.

Mais pour beaucoup de personnes, leur emploi donne du sens à leur vie. Ce qui signifie que se retrouver sur le banc de touche peut avoir des conséquences très 91

négatives. Se retrouver en quelque sorte perdu, sans savoir quoi faire. Nous sommes des mammifères, mais à la différence des autres, nous avons la capacité de fabriquer des constructions mentales qui sont des abstractions par rapport au réel. Tout ce qui a de l'importance dans notre vie est au finale une construction mentale. Les nations ne sont que des lignes imaginaires tracées sur une carte, c'est à dire des constructions mentales. L'argent n'a pas d'existence concrète, c'est simplement un moyen commun pour

commercer, donc une construction mentale. La Foie, la religion, la spiritualité sont des systèmes de croyances, autrement dit des constructions mentales. Et même la notion de moi, le ”je”, notre propre identité, ne pourrait être qu’une construction mentale. En tout cas c’est ce qui est au cœur de la philosophie bouddhiste.

Ces constructions mentales sont comme des illusions très difficiles à percer.

Beaucoup nous accompagnent depuis notre naissance. Mais il y en a une que l’on doit casser pour changer les fondements de la société. Il s’agit de la construction mentale de l’emploi.

D’ailleurs, le mot “emploi” est très récent au final. Le terme anglais "Job" est apparu dans le lexique au début de la révolution industrielle il y 250 ans. Nous pensons depuis notre enfance que “emploi” = “travail”, que c’est la même chose au final. C’est faux ! C’est une construction mentale. Un emploi est une sous catégorie du travail. Un emploi n’a qu’une seule fonction : nous apporter un revenu. Donc emploi = argent. Pas travail. Lorsqu’on perd un emploi, soit on en cherche un autre similaire dans le même domaine d’activité, soit on change de branche en faisant une formation pour trouver un emploi différent. Le travail c’est quelque chose que l’on se créer pour nous même. C’est quelque chose qui nous donne véritablement du sens. Ça peut être travailler sur un projet personnel qui vous tient à cœur, ou pour la communauté, chacun s’approprie cette notion de travail à sa guise. Si vous confondez emploi et travail, alors lorsque les robots prendront votre emploi, votre “job”, vous allez vous sentir extrêmement déprimé, car ce sera comme si vous aviez perdu le sens de votre vie. Alors qu’en réalité, vous aurez simplement perdu une illusion. Une construction mentale.

Pour résumer le plus simplement possible, posez vous cette question : *"Est-ce que je me lèverais tous les matins, 5j/7 pour faire mon boulot, 40h/semaine, si en retour, je n’étais pas payé ?”* . Si la réponse est non, c’est que ce boulot est un emploi. Si vous pensez que oui, alors c’est un travail.

C’est ce changement de perspective que nous devons adopter à l’échelle sociétale.

Et c’est loin d’être impossible puisque certains le font déjà. Ce sont les personnes qui se sont créé leur propre sens, leur propre travail. C’est le professeur de littérature qui a lancé une plateforme de cour en ligne, c’est le sportif, le photographe, le comédien, le 92

youtubeur, le blogueur, et toutes ses personnes qui se sont créé leur propre travail et qui réussissent à en vivre. C’est là où se trouvent nos rêves, nos aspirations. Le futur du travail, ce n’est pas d’aller chercher un emploi en espérant que les entreprises en ont encore à disposition. Le futur du travail, c’est ce que nous allons créer pour nous même et qui fera sens.

Mais la question légitime qui se pose ensuite c’est : “Si je ne travaille plus pour gagner de l’argent, comment vais je subvenir au besoin de base ?”. Et c’est là où l’automatisation entre dans le débat politique. Si les entreprises automatisent de plus en plus d’emploi dans un très grand nombre de disciplines, augmentant drastiquement leur profit, ont-elles une responsabilité vis-à-vis des emplois perdus ? Est ce que les gouvernements doivent réfléchir à un meilleur système de redistribution des richesses

? La société ne pourra pas fonctionner sur le même modèle si les machines sont meilleures et plus productives dans presque toutes les disciplines. Cela fait partie des grands challenges du 21e siècle. L’historien Yuval Noah Harari voit arriver une nouvelle classe sociale due à l’automatisation : la classe des inutiles. Comme mentionné précédemment, les humains ont deux sortes de capacités qui nous rendent utiles sur le marché du travail : physiques et cognitives. La révolution industrielle a peut-être eu pour résultat des machines qui ont éliminé les humains dans des emplois nécessitant de la force et des actions répétitives. Mais avec nos capacités cognitives, les humains étaient largement en sécurité dans leur travail. Pour combien de temps encore ? Les IA commencent maintenant à surpasser les humains dans certains domaines cognitifs. Et bien que de nouveaux types d’emplois apparaîtront certainement, nous ne pouvons pas être sûrs selon Harari, que les humains les feront mieux que les IA. Au final, elles n'ont pas besoin d'être plus intelligentes que les humains pour transformer le marché du travail. Elles ont juste besoin de savoir faire la tâche plus rapidement et efficacement.

Malgré tout, les hommes sans emploi ne sont pas des êtres humains inutiles.

Les personnes sans emplois sont toujours valorisées. Mais Harari a une définition spécifique lorsqu'il parle de classe inutile. Il s'agit d'inutile du point de vue du système économique et politique, et non d'un point de vue moral. Les structures politiques et économiques modernes ont été construites sur des êtres humains utiles à l'État, notamment en tant que travailleurs et soldats. Si ces rôles sont assumés par les machines, nos systèmes politiques et économiques cesseront tout simplement d'attacher beaucoup de valeur aux humains.

Que devrions-nous faire ? Selon Harari, il faut l'intégrer à l'agenda politique, pas seulement à l'agenda scientifique. C'est quelque chose qui ne devrait pas être laissé aux chercheurs et sociétés privées. Ils en savent beaucoup sur les techniques, l'ingénierie, mais ils n'ont pas nécessairement la vision et la légitimité nécessaires pour décider de l'avenir de l'humanité.

93

Alors ça sonne un peu comme la fin du monde, ou en tout cas, comme quelque chose de très négatif. En réalité, nous ne devons pas en avoir peur. Il faut plutôt en parler. Discuter et être conscient que c'est un train qui a déjà quitté la station et qui sera bientôt là. Si l'on regarde les 150 dernières années, on remarque que le niveau de vie moyen planétaire s'est vraiment amélioré. Nous vivons la meilleure époque dans l'histoire de l'humanité. Même si nous avons du mal à le réaliser. Donc cette tendance à travailler moins et mieux vivre ne date pas d'hier et est apparue, car nous avons adopté la technologie au cœur de nos sociétés. Cette tendance continue de manière exponentielle et l'arrivée de l'intelligence artificielle marque un tournant majeur.

Si rien ne change dans nos sociétés, alors oui, de nombreux problèmes vont émerger. Le taux de chômage va exploser, la fracture sociale également. Nous devons réfléchir à un nouveau système. Un système qui s'oriente sur le fait que nous n'avons plus à travailler. Ce n'est plus OBLIGATOIRE pour être un citoyen et bien vivre. Une société future où on travaille seulement si on le souhaite. Pour faire ce qu'on aime, exercer nos passions. Si tous les travaux pénibles qui nous minent le moral tous les matins sont exécutés par des machines, alors on pourra se lever avec une motivation

débordante pour travailler dans ce que l'on aime. Et dans cette société future, nous n'aurons pas à travailler pour gagner de l'argent. Travailler aura un sens bien plus profond. Il est facile de dresser un portrait négatif de l'automatisation des emplois, car cela entraîne le licenciement de millions de personnes . Mais ce dont on parle moins, c'est qu'un très grand pourcentage des emplois automatisés sont des tâches déshumanisantes que personne n'a réellement envie de faire. Miner du charbon dans une mine insalubre, conduire un camion toute la journée, passer des heures à récolter des oignons à coup de bêche, scanner des articles de manière robotique dans un supermarché, ce n'est pas le genre de métier qui fait rêver les enfants lorsqu'on leur demande ce qu'ils veulent faire lorsqu'ils seront plus grands. Le bon côté de l'automatisation c'est affranchir l'humanité des tâches qui nous rendent esclaves du monde du travail. Les slogans des syndicats ne doivent pas être : "Sauvons nos emplois !!". Ce n'est pas les emplois qui importent, mais bien les humains derrière.

Nous devons réfléchir à la transition entre une société où l'on doit travailler pour survivre, à une société où travailler n'est qu'optionnel. Où un travail n'est pas nécessaire pour bien vivre.

Une approche pour compenser le bouleversement social à venir est le "revenu de base universel". Pour faire simple, l'idée c'est que s'il n'y a pas assez d'emplois, le gouvernement ou une autre entité paieront à chacun une somme minimum pour subvenir aux besoins de base comme la nourriture, le logement, la santé ou l'éducation. Mais certaines personnes, comme l'historien Yuval Noah Harari, craignent que le revenu de base ne soit pas "universel". C'est-à-dire que les pays riches pourront se le permettre, mais pas les pays pauvres. Le résultat serait alors un fossé toujours plus grand entre les pays riches et les pays pauvres.



Le problème étant que les gouvernements sont les principaux fournisseurs d'un revenu de base. Cela fonctionnera bien pour les États-Unis, ou les pays européens, qui peuvent taxer les entreprises à revenu élevé telles que Google et Facebook. Mais ces taxes n'aideront pas les habitants de pays moins développés, comme le Honduras ou le Bangladesh, qui seront mis

au chômage par une automatisation croissante. Et il est naïf d'imaginer que les taxes sur Google iront aux habitants du Bangladesh. La révolution de l'automatisation rendra probablement certaines régions du monde extrêmement riches et puissantes tout en détruisant complètement l'économie des autres.

Tous les experts ne sont toutefois pas pessimistes quant au revenu de base. Il est possible que l'automatisation généralisée conduise à une ère d'abondance mondiale.

Encore faut-il trouver comment partager les richesses.

**Créativité**

Si de plus en plus d'activité humaine sont maîtrisées voir surpassées par des intelligences artificielles, assurément, ce ne sera jamais le cas de la musique, la peinture, le cinéma, et la science ! En effet, la créativité est considérée comme l'un des piliers de ce qui définit l'humanité. Même si de nombreuses espèces animales créent des constructions époustouflantes, pensez à la toile délicate d'une araignée, elles sont généralement créées dans un but pratique, comme attraper une proie ou séduire un partenaire.

Les humains, cependant, font de l'art pour eux-mêmes, en tant que forme d'expression personnelle. Donc est-ce qu'une intelligence artificielle peut être créative

? Une chose est sur avec l'intelligence artificielle c'est qu'il ne faut pas sous-estimer ses possibles capacités futures.

Il existe un champ d'études dédié aux développements de système artificiel créatif qui s'appelle la créativité informatique (Computational creativity). Le but de cette approche multidisciplinaire est de modéliser, simuler ou reproduire la créativité à l'aide de l'informatique, pour atteindre l'un des objectifs suivants:

- Construire un programme ou un ordinateur capable de créativité humaine.

- Mieux comprendre la créativité humaine et formuler une perspective algorithmique sur le comportement créatif de l'homme.

- Concevoir des programmes pouvant renforcer la créativité humaine sans être nécessairement eux-mêmes créatifs.

Mais le domaine est toujours entravé par un certain nombre de problèmes fondamentaux. La créativité est très difficile, voire impossible, à définir en termes objectifs. Est-ce un état d'esprit, un talent, une manifestation génétique, une capacité, 95

ou un processus ?

Ce sont des problèmes qui compliquent l'étude de la créativité en général, mais certains problèmes se rattachent spécifiquement à la créativité informatique:

- La créativité peut-elle être programmée ? Dans les systèmes existants auxquels la créativité est attribuée, la créativité est-elle celle du système ou celle du programmeur ?

- Comment évaluons-nous la créativité informatique ? Qu'est-ce qui distingue la recherche en créativité informatique de la recherche en intelligence artificielle en général ?

- Si la créativité consiste à enfreindre les règles ou à renier les conventions, comment un système algorithmique peut-il être créatif ? Si une machine ne peut faire que ce pour quoi elle a été programmée, comment son comportement peut-il être qualifié de créatif ?

En effet, tous les théoriciens de l'informatique n'acceptent pas le principe selon lequel les ordinateurs ne peuvent faire que ce pour quoi ils ont été programmés. Un point clé en faveur de la créativité informatique.

Lorsque votre morceau préféré passe à la radio, que ressentez-vous ? Vous êtes plutôt content, satisfait, voire même euphorique n'est-ce pas ? Maintenant, si vous appreniez que ce même morceau a été composé par

un ordinateur, un programme, une intelligence artificielle. Est-ce que ça change votre ressenti de cette musique ?

En 1843, Ada Lovelace (1815 - 1852), une mathématicienne anglaise considérée comme la première programmeuse informatique, a écrit que les ordinateurs ne pourront jamais avoir une intelligence comparable à l'humain tant qu'ils suivront uniquement ce que les programmeurs leur ont demandé de faire. Selon elle, une machine doit être capable de créer des idées originales pour être considérée comme intelligente. Le test de Lovelace, formulé en 2001, propose un moyen d'évaluer cette idée. Un ordinateur réussit le test s'il arrive à produire quelque chose que ses programmeurs n'arrivent pas à expliquer en regardant uniquement son code. Il s'agit d'un test de pensée plutôt que d'un test scientifique, mais c'est un bon début dans cette recherche pour détecter l'étincelle créatrice des machines.

À première vue, l'idée d'un ordinateur créant une musique originale de très haute qualité peut paraître impossible. On peut très bien créer un algorithme qui est capable de générer de la musique en utilisant des nombres aléatoires, des fonctions chaotiques, etc. Le résultat est une séquence de notes musicales possédant de nombreux passages extrêmement originaux, mais seulement une fraction est supportable à l'oreille humaine. Ainsi, l'ordinateur n'a aucun moyen de distinguer les passages que l'on 96

considère magnifiques de ceux qui nous percent les tympans.

Par contre, il y a aussi la possibilité de regarder des processus naturels où la créativité peut émerger : l'évolution. En effet, la nature, à travers l'évolution, a abouti à des résultats que l'on considère magnifiques, beaux, créatifs. Les motifs des ails des papillons, le chant des oiseaux, les pétales de fleurs, etc. Du coup, une approche très prometteuse pour que les machines puissent aboutir à des résultats originaux est ce qu'on appelle les "algorithmes évolutionnaires.

Donc si on reprend l'exemple de la musique. Comment l'évolution peut-elle permettre aux machines de devenir musicalement créative ? Tout d'abord, on peut introduire dans la machine une "phrase musicale" de 10 secondes par exemple.

Ensuite, on insère un algorithme basique qui mimique et opère des mutations aléatoires afin de créer des phrases musicales différentes en s'inspirant, changeant, réordonnant, combinant les notes de celle de base. Maintenant que nous avons une nouvelle génération de phrase musicale, on peut appliquer une fonction de sélection.

Tout comme la sélection naturelle est déterminée par l'environnement extérieur, la fonction de sélection peut être déterminée par une mélodie extérieure choisie par des musiciens humains ou des fans de musique. Le but est de définir ce qui est considéré comme une belle mélodie pour un grand nombre de sujets.

Une fois cette mélodie sélectionnée, l'ordinateur peut comparer toutes ses phrases musicales avec la mélodie jugée "belle" et peaufiner sa composition. Une fois tous les passages les moins similaires enlevés, l'algorithme peut à nouveau appliquer des combinaisons et mutations sur ce qui reste pour générer de nouvelles phrases musicales. Et à nouveau comparé avec la mélodie "belle" via la fonction de sélection.

Ce processus peut être répété en boucle jusqu'au résultat final.

Cette approche a tellement de complexité et de caractère aléatoire que le résultat pourrait réussir le test de Lovelace. Mais le plus important c'est que vu l'apport extérieur du facteur humain dans le processus, le résultat pourrait être jugé magnifique selon nos propres critères.

Dans le domaine de la musique classique contemporaine, Iamus est le premier ordinateur à composer à ne partir de rien et à produire les partitions que les interprètes professionnels peuvent jouer. Le London Symphony Orchestra a joué une pièce pour orchestre complet, incluse dans le premier CD d'Iamus, que le journal "New Scientist"

a décrit comme "la première œuvre majeure composée par un ordinateur et interprétée par un orchestre". La mélomique, la technologie derrière Iamus, est capable de générer des morceaux de différents styles de musique avec un niveau de qualité similaire.

La recherche sur la créativité dans le jazz s'est concentrée sur le processus d'improvisation et les exigences cognitives que cela impose. Le robot Shimon, développé par Gil Weinberg de Georgia Tech, a fait preuve d'improvisation sur du jazz. Les logiciels d'improvisation virtuelle basés sur des modèles d'apprentissage automatique de styles musicaux, tels que OMax, SoMax et PyOracle, sont utilisés pour créer des improvisations en temps réel en réinjectant des séquences de longueur variable apprises à la volée par un artiste.

La créativité informatique dans les arts visuels a connu des succès notables dans la création d'art abstrait. Le programme le plus célèbre dans ce domaine est Aaron, de Harold Cohen, qui est développé depuis 1973. Aaron génère des dessins en noir et blanc ou des peintures en couleur incorporant des figures humaines ( comme des danseurs), des plantes, des roches et d'autres éléments. Ces images sont d'une qualité suffisamment élevée pour être affichées dans des galeries de bonne réputation.

Parmi les autres artistes logiciels connus, citons le système NEvAr (pour "Art neuro-évolutionnaire") de Penousal Machado. NEvAr utilise un algorithme génétique pour dériver une fonction mathématique qui est ensuite utilisée pour générer une surface tridimensionnelle colorée. Un utilisateur humain est autorisé à sélectionner les meilleures images après chaque phase de l'algorithme génétique. Ces préférences sont utilisées pour produire les surfaces les plus attrayantes pour l'utilisateur.

Un domaine émergent de la créativité informatique est celui des jeux vidéo.

ANGELINA est un système de développement créatif de jeux vidéo en Java. Un aspect important est Mechanic Miner, un système capable de générer de courts segments de code agissant comme une simple mécanique de jeu. ANGELINA peut évaluer l'utilité de ces mécanismes en jouant et en vérifiant si le nouveau mécanisme permet de créer un jeu intéressant.

Mais est-ce que cela satisfait vraiment notre intuition de ce qui est créatif ou non ?

Est-ce suffisant pour créer quelque chose de beau et original ? Ou est-ce que la créativité nécessite une intention et une conscience de ce qui est en train d'être créé ?

Et au final, qu'est-ce que la créativité humaine ? Est-ce quelque chose qui dépasse juste un assemblage de neurones connecté entre eux qui s'agence en fonction de nos expériences ?

Sculpture, poésie, musique, peinture, littérature, tous ses domaines artistiques sont-ils le propre de l'Homme ? Les resteront-ils ? Pour beaucoup d'entre nous, les créations artistiques des machines ne sont encore pas au niveau pour être jugées créatives et artistiques. Mais cela n'empêche pas certains collectionneurs d'acheter des peintures créées par une IA. En octobre 2018, le tableau "Portrait d'Edmond de 98

Belamy" a été vendu pour $432 000 lors d'une enchère. De quoi peut être lancer un nouveau mouvement artistique ?

Mais si une oeuvre artistique vous fait vibrer d'émotion, pleurer, vous impressionne ou encore vous donne la chair de poule, est-ce que cela importe vraiment de savoir qui ou quoi l'a créée ?

**Médecine**

Selon Wikipedia, la médecine (du latin : medicina, qui signifie "art de guérir, remède, potion") est la science et la pratique étudiant l'organisation du corps humain, son fonctionnement physiologique, et cherchant à préserver la santé (physique et mentale) par la prévention et le traitement des pathologies.

Une chose est sûre, la médecine a toujours été entre les mains de personne qualifiée, spécialisée et bien sûr … humaine. Comme on l'a vu, le monde s'automatise. Mais à mesure que l'intelligence artificielle étend ses tentacules à travers les différents secteurs d'activité, de plus en plus de personnes se demandent s'il y a des tâches et des métiers qui resteront toujours uniques à l'être humain. Car lorsqu'on pense aux machines sur le marché du travail, on visualise les lignes d'assemblage dans les usines automobiles ou encore les métiers liés à la conduite de

véhicule (Taxi, ambulancier, routier, etc.). En revanche, on a tendance à se dire que les métiers créatifs ou qui requièrent un haut niveau d'étude et de capacité cognitives resteront dans les mains d'Homo Sapiens.

En ce qui concerne la médecine, le métier que l'on a tout de suite en tête c'est celui de docteur. Le médecin généraliste que l'on va voir dès qu'on a un petit souci et qui, après avoir attendu des heures dans la salle d'attente à attraper les virus des autres, nous prescrit les médicaments dont on a tant besoin. Assurément, les docteurs ne peuvent pas être remplacés ! Eux qui sont connus pour avoir fait de longues études.

Et bien si on y regarde de plus près, la première et principale fonction d'un docteur c'est de correctement diagnostiquer une maladie et ensuite, suggérer le meilleur traitement disponible. Si je vais voir mon généraliste en me plaignant de fièvre et diarrhée, ces symptômes peuvent résulter d'un virus intestinal, du choléra, la malaria, le cancer ou même une maladie inconnue, qui sait. Mon docteur n'a que quelques minutes pour faire le bon choix, car il a d'autres patients derrière moi. Il faut bien qu'il gagne sa vie ! Après quelques questions, il met en parallèle mon dossier médical et les maladies humaines qu'il connaît, et finalement, me dit que je dois avoir mangé un truc pas frais et me prescrit un médicament. Aussi bon soit-il, mon docteur ne peut pas savoir tous les aliments que j'ai ingérés au cours des derniers jours. Il ne peut pas être au courant de toutes les dernières découvertes en médecine, les derniers traitements, ni avoir lu les nouvelles publications. En plus, mon docteur est parfois fatigué voir lui même malade ce qui peut affecter son jugement. Pas étonnant donc qu'il lui arrive de se tromper dans ses diagnostics en me prodiguant le traitement le 99

moins efficace.

Maintenant, faisons un bon dans le futur proche. Je me retrouve donc à nouveau avec fièvre et diarrhée. Seulement cette fois-ci, je ne vais pas me rendre dans un cabinet ou à l'hôpital. Je vais juste consulter mon intelligence artificielle personnelle qui se trouve chez moi. Comparé à mon docteur de 2019, mon IA possède beaucoup d'avantages. Déjà, elle peut se connecter à une immense base de données contenant toutes les maladies connues dans l'histoire de l'humanité. Elle peut aussi mettre à jour cette base

de données à la minute même où un nouveau traitement est publié, mais également avec les statistiques de tous les autres patients sur la planète. Ensuite, mon IA me connaît des pieds à la tête, mais également de mes cellules à mon ADN. Elle possède mon historique médical, celui de mes parents, frères et soeurs, cousins, voisins et amis. Mon IA saura si j'ai récemment visité un pays tropical, ce que j'ai mangé ces derniers jours au nutriment près, si j'ai régulièrement des intoxications alimentaires, si des membres de ma famille ont déjà eu des cancers de l'estomac et s'il y a d'autres personnes souffrant des mêmes symptômes dans le voisinage. Tout cela, en une fraction de seconde. Et enfin, mon IA ne sera jamais fatiguée ou malade et pourra se consacrer uniquement sur mon cas pendant des heures s'il le faut, sans avoir à réduire la séance car elle a d'autres patients après moi. Et, elle sera disponible même si je me trouve en vacance en Nouvelle-Zélande, 24h sur 24, 7j sur 7.

Il y a donc fort à parier que son diagnostic et sa recommandation du traitement optimal seront bien plus précis que ceux de mon docteur en 2019. Et ce n'est pas de la Science Fiction. Le fameux Watson, d'IBM, après avoir gagné le jeu TV Jeopardy, a désormais pour mission de diagnostiquer des maladies. Si vous avez prévu d'entrer en fac de médecine pour devenir médecin généraliste à la fin des années 2020 … vous ne devez pas voir cela d'un très bon oeil et peut-être qu'il n'y aura tout simplement aucun travail pour vous. Ou alors, vous serez plutôt une sorte d'informaticien/docteur et l'IA fera 80% du boulot !

Watson d'IBM est destiné au diagnostic médical ©Raysonho CC BY 3.0

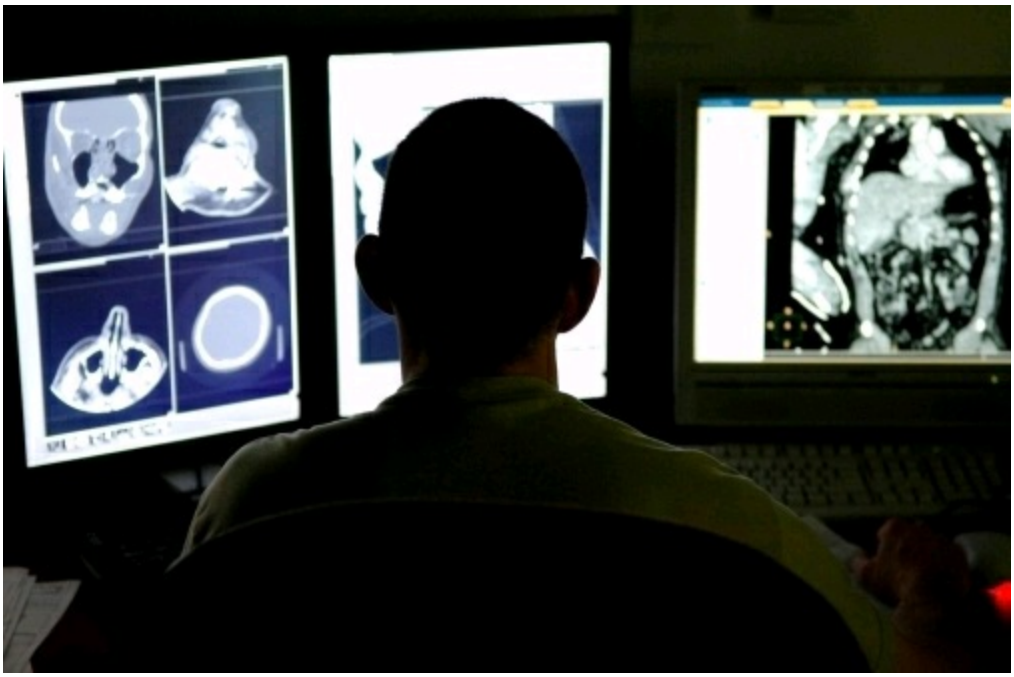
Ce genre d'IA sera également présente dans les domaines médicaux spécialisés comme l'oncologie. En effet, lors d'un test, un algorithme a diagnostiqué correctement 90% des cancers des poumons qui lui étaient présentés, alors que les docteurs ont obtenu un ratio de 50%. Aujourd'hui, des algorithmes sont utilisés lors de scanner et mammographie ce qui donne un deuxième avis et parfois, détectent des tumeurs que les médecins ont loupées. Récemment, une équipe de chercheurs à Oxford ont développé Ultromics, une IA spécialisée dans le diagnostic de maladie cardio-vasculaire. En écoutant les battements cardiaques d'un patient, elle est capable de déterminer si ce dernier souffre ou pas d'un problème au coeur.

Les algorithmes de deep learning ont été en mesure de diagnostiquer la présence ou l'absence de tuberculose sur des radiographies du thorax avec une précision étonnante. Les chercheurs ont d'abord entraîné les IA avec des centaines de radios de patients sans et avec tuberculose. Ensuite, ils ont testé ces IA avec 150 nouvelles radiographies. Les algorithmes

ont atteint un taux de précision impressionnant de 96%

- meilleur que celui de nombreux radiologues humains - et les chercheurs pensent qu'ils peuvent améliorer ce taux avec davantage de formation et de deep learning plus avancés. La détection automatisée de la tuberculose sur des radiographies pulmonaires peut faciliter les efforts de dépistage et d'évaluation dans les zones où les cas de tuberculose sont élevés, mais l'accès aux radiologistes limité.

101



Des IA sont désormais plus performantes pour repérer des anomalies sur des

radios.

Des algorithmes de deep learning similaires ont montré des succès encourageants dans d'autres branches de la médecine telles que l'ophtalmologie et la cardiologie. Des chercheurs de Google ont pu former une IA à détecter la propagation du cancer du sein dans les tissus des ganglions lymphatiques sur des images microscopiques avec une précision comparable (ou supérieure) aux pathologistes humains. Il est difficile de chercher de minuscules cellules cancéreuses. C'est comparable à trouver

une seule maison sur une photographie satellite d'une ville entière. Un humain peut souffrir de fatigue ou d'inattention en analysant les images, tandis qu'une IA peut traiter des millions de pixels sans transpirer.

La chirurgie est incroyablement complexe, nécessite une intense formation spécialisée et est, littéralement, une question de vie ou de mort. L'intelligence artificielle pourrait aider à réduire les risques. Lorsqu'elle est associée à des programmes de réalité augmentée, une IA peut fournir aux chirurgiens des informations en temps réel et surimposé sur le corps du patient. Pour certaines chirurgies, une IA pourrait contrôler un bras robotique et opérer elle même le patient.

Il reste bien sûr pas mal de problèmes à résoudre avant de voir des IA rendre les docteurs obsolètes. Mais peu importe les difficultés, il nous faudra les résoudre qu'une seule fois. Alors que pour former un seul médecin humain, cela coûte beaucoup d'argent et prend des années. Juste pour UN médecin. Si vous en voulez deux, il faut répéter le processus. Alors que pour des IA, une fois que ce sera au point, il n'y en aura pas qu'une seule, mais des millions et même si cela coûte 100 milliards 102

à mettre au point, sur le long terme, c'est bien plus avantageux que de former des médecins humains. Et une telle technologie offre le meilleur système de santé jamais conçu et applicable dans le monde entier.

Mais même si une IA devient techniquement meilleure qu'un médecin, il reste néanmoins la touche humaine. Si un scanner indique un jour que vous avez un cancer, est-ce que vous voulez recevoir la nouvelle de la part d'une machine, ou par un médecin attentionné et empathique ? Et pourquoi pas d'une machine capable de compassion, d'empathie et qui adapte chaque mot à votre personnalité, car elle vous connaît mieux que vous ne vous connaissez vous même ? Est-ce que le soutien émotionnel est réservé aux humains ? On pourrait penser que oui, mais encore une fois, c'est très probable que non. Les émotions ne sont pas un truc métaphysique que Dieu nous a donné pour apprécier la poésie. Ce n'est ni plus ni moins qu'une série de données biochimique présent dans le règne animal, pas seulement chez Homo Sapiens. Et qu'est-ce qui est performant à collecter des données ? Les IA ! Donc une machine pourrait très bien non seulement diagnostiquer parfaitement votre problème, vous donnez

le traitement optimal, mais également lire vos émotions et réagir en conséquence. Tout comme le fait un docteur au final. Après tout, comment un médecin fait-il pour connaître votre condition émotionnelle ? Il se fie à vos signaux extérieurs : Expression faciale, langage corporel. Et également à ce que vous lui dite et comment vous le dites. Tout cela est tout à fait faisable par un système artificiel.

La médecine liée à l'intelligence artificielle sera également orientée vers la prévention. Car l'un des avantages essentiels de l'IA provient de sa capacité à rassembler et à analyser des quantités de données et à tirer des conclusions de son analyse. Qui est le plus susceptible d'avoir un cancer ? Quels sont les facteurs de risque qui rendent un patient plus vulnérable, par exemple, aux crises cardiaques, par opposition aux accidents vasculaires cérébraux ?

On voit déjà la confiance placée dans les algorithmes se manifester dans le domaine de la médecine. Les plus importantes décisions concernant votre santé ne seront pas prises par vous sur la base de vos émotions et sensations, mais par des algorithmes, sur la base de ce qu'ils savent sur vous. Si cela paraît flou et théorique, voici un exemple pratique :

Le 14 mai 2013, l'actrice Angelina Jolie a publié un article dans le New York Times sur sa décision de subir une double mastectomie. Ce n'est pas une petite décision !!! Angelina Jolie a vécu pendant des années sous l'ombre du cancer du sein.

Sa mère et sa grand-mère en étant mortes à un âge relativement précoce. Angelina Jolie a fait un test génétique qui a prouvé qu'elle portait une mutation dangereuse du gène BRCA1. Selon des récentes statistiques, les femmes porteuses de cette mutation ont une probabilité de 87% de développer un cancer du sein. Même si, à l'époque, elle n'avait pas de cancer, elle a décidé d'anticiper la maladie et d'avoir une double mastectomie. Quand Angelina Jolie a dû prendre une des plus importantes décisions 103

de sa vie, elle n'est pas allée dans la forêt pour tenter de se connecter à ses sentiments les plus intimes via 10h de méditation. Au lieu de cela, elle

a préféré écouter ses gènes, dont la voix s'est manifestée non pas par des émotions, mais par des chiffres.

Elle a préféré écouter un algorithme plutôt que ses propres sensations.

Il y a plusieurs années, Google a créé son étude Baseline, une entreprise complète et ambitieuse impliquant des milliers de bénévoles et 100 spécialistes dans différents domaines médicaux. Comme son nom l'indique, le but de l'étude était d'établir une sorte de base de la santé humaine à partir de laquelle des algorithmes et des chercheurs pourraient isoler des indices biologiques susceptibles d'être des prédispositions à des maladies spécifiques.

Aujourd'hui, l'étude Baseline se poursuit sous la bannière de Verily, une division d'Alphabet (la société mère de Google). Dans un avenir proche, il est facile d'imaginer un monde où les maladies non transmissibles (accidents vasculaires cérébraux, cancers, crises cardiaques) ou les maladies héréditaires sont identifiées lors d'une seule visite chez le médecin. Les patients pourront non seulement voir la probabilité qu'ils aient de contracter une maladie spécifique, mais les médecins pourront également les aider à prévenir ces conditions grâce à un plan d'action clair et sur le long terme.

Aujourd'hui, vous pouvez très bien vivre des années avec des cellules cancéreuses qui se multiplient. Malheureusement, le jour où vous vous en rendez compte, c'est peut-être trop tard. Bientôt, vous irez aux toilettes pour faire la grosse commission.

Vos toilettes intelligentes connectées vous enverront ensuite une notification sur votre téléphone : "Bonjour, après avoir analysé vos excréments, j'ai remarqué 100 000

cellules cancéreuses en provenance des poumons. Il est recommandé d'arrêter de fumer ou bien vous développerez un cancer dans 25 ans. La probabilité est de 89%".

En regardant votre téléphone, vous saurez en permanence votre pression sanguine, votre taux de cholestérol, la quantité de nutriment que vous avez

ingéré lors de votre déjeuner et si vous avez des carences en fer ou protéine. En faite, on pourrait même être constamment contrôlé par des systèmes externes intelligents dans notre quotidien sans vraiment y prêter attention. En prenant une douche, vous passerez en même temps un scanner pour vérifier les traces de cancer de la peau. Chaque fois que vous saisirez votre téléphone, vous obtiendrez une électrocardiographie, chaque fois que vous vous regarderez dans le miroir, vous aurez un examen rétinal, chaque fois que vous dormirez, votre matelas analysera votre respiration, et pression sanguine. De nombreux systèmes intelligents peuvent être incorporé dans notre vie de tous les jours, tout autour de nous afin d'être en écoute constante pour prévenir d'éventuels problèmes de santé. Et lorsque nous avons en permanence des diagnostics sur l'évolution de notre santé, les chances de traiter les pathologies sont extrêmement plus élevées.



Toutes ces données collectées joueront un rôle crucial dans la médecine préventive du futur. Ce qui est appelé le "big data" sera capable de prévenir une maladie pour un individu, mais également une menace d'épidémie pour une société tout entière.

Aujourd'hui, il faut plusieurs jours pour que le ministère de la Santé réalise qu'une épidémie de grippe est en train de se répandre dans le pays. Google pourrait le faire en une minute. Tout ce dont l'algorithme a besoin, c'est de surveiller certains mots clés qui sont tapés dans les emails et le moteur de recherche des Français au cours d'une minute. Ensuite, l'algorithme analyse les mots clés : Maux de tête, toux, fièvre, éternuement. Admettons qu'en moyenne, ils sont répertoriés 100 000 fois. Mais si, lors de cette fameuse minute, ils sont répertoriés 500 000 fois, alors Google peut conclure qu'une épidémie de grippe a éclaté. Les autorités peuvent ensuite réagir plus vite. Mais pour que Google réussisse ce tour de magie digne de minority report, il faut qu'on l'autorise à lire nos informations et les partager avec le ministère de la Santé.

Sacrifier notre vie privée pour augmenter le niveau de santé des citoyens. Voilà peut-

être le choix qui nous sera demandé.

Mais ces algorithmes seront si bons à prendre des décisions pour nous que ce sera considéré comme de la folie de ne pas suivre leur recommandation. Il est important de noter que si l'intelligence artificielle révolutionne certainement nos relations avec la médecine, elle est beaucoup plus susceptible de le faire de manière subtile et discrète.

Les soins de santé deviendront plus précis, plus complets et moins chers avec le temps, ce qui est une bonne nouvelle pour tout le monde.

**Assistants personnels**

En termes d'assistant personnel, la science-fiction nous a habitués à voir les héros évoluer conjointement avec des intelligences artificielles ayant pour but de les aider.

On pense à Jarvis dans Iron Man, ou encore Samantha dans le film "Her".

Aujourd'hui, regardez n'importe quelle chaîne de télévision et vous tomberez surement sur un spot pour Amazon Echo et Google Home. Telle est la rapidité avec laquelle ces produits évoluent. Les experts prédisent que Siri, Alexa, Cortana et Google Assistant se répandront dans les foyers à travers smartphones et produits connectés, même s'ils ont un long chemin à parcourir avant d'atteindre une véritable intelligence. Ceci dit, ce n'est pas non plus ce qu'on leur demande. En tous cas, pas dans l'immédiat, et ce n'est certainement pas le fait que ces IA ne sont pas plus intelligentes qu'un enfant de 2 ans qui ralentira leur progression sur le marché.

Siri d'Apple, reçoit l'honneur d'être la première IA personnelle, lancée sur l'iPhone en 2011. À l'époque, Siri était révolutionnaire et a eu le monopole sur le marché pendant plusieurs années, jusqu'à ce que Amazon lance Alexa et que Microsoft fasse 105

ses débuts avec Cortana en 2014, Google Assistant est arrivé en 2016, Samsung a rejoint la fête avec Bixby en 2017.

Alexa et Google Assistant ont pris beaucoup de place ces dernières années, grâce notamment à un marketing agressif. Jusqu'à présent, les efforts ont été placés dans la reconnaissance vocale afin de savoir ce que l'utilisateur dit. La prochaine étape s'agira davantage de savoir pourquoi et où il le dit. La compréhension contextuelle est la clé pour que la recherche vocale puisse devenir une partie intégrante de la vie des consommateurs.

L'assistant Google a pris les devants, car la technologie peut mieux comprendre l'importance du contexte lorsqu'une question est posée. Les utilisateurs peuvent demander à Google Home : "*Quelle est la météo pour demain?* ". Puis demander :

"*Est-ce qu'il va neiger ?* ". Google Home sait que la deuxième question concerne leur emplacement spécifique. Alexa, en revanche, ne comprend pas le contexte dans ce même cas. Google Assistant peut comprendre les demandes d'utilisateurs spécifiques et personnaliser la réponse, ce qui le place un cran au-dessus de la compétition. Les assistants vocaux continueront à offrir des expériences plus personnalisées au fur et à mesure qu'ils parviendront à mieux différencier les voix. En 2017, Google a annoncé que son assistant Google Home pouvait prendre en charge jusqu'à six comptes d'utilisateur et détecter des voix uniques, ce qui permet aux utilisateurs de Google Home de personnaliser de nombreuses fonctionnalités. Une personne peut demander

" *Que reste-t-il à faire sur mon calendrier aujourd'hui ?* ". Et l'assistant reconnaîtra la voix et saura quel calendrier décrire. Il inclut également des fonctionnalités telles que les pseudonymes, les lieux de travail, les informations de paiement et les comptes associés tels que Google Play, Spotify et Netflix. Amazon se bat pour rattraper son retard et la société a annoncé la détection de voix personnalisée pour des expériences plus personnalisées il y a quelques mois à peine.

Le futur des assistants virtuels inclut également la mobilité. Nous aurons en permanence notre assistant virtuel sur nous. C'est déjà le cas en quelque sorte puisque Siri et Google assistant se trouvent sur nos téléphones, mais il faut s'attendre à une interaction encore plus importante avec ces derniers. L'utilisation des oreillettes va se généraliser et nous aurons tendance à communiquer plus souvent de manière vocale

avec notre téléphone. Ce qui fait que nous le sortirons moins de notre poche. Voir des personnes parler toutes seules dans la rue ne sera plus associé à la folie mais simplement à une interaction homme/machine. L'IA qui sera dans notre poche sera la même qui se trouve dans notre maison, la même qui se trouve dans notre voiture et la même qui se trouve dans notre bureau.

Lors d'une conférence début 2018, Google a présenté sa dernière mise à jour 106

concernant son assistant virtuel. Nommée "Duplex", l'intelligence artificielle est capable de prendre un rendez-vous téléphonique au nom de son utilisateur. Durant la présentation, l'exemple concernait un rendez-vous chez le coiffeur. Une fois l'instruction donnée, l'intelligence artificielle compose le numéro du salon de coiffure et entame une conversation en temps réel avec la personne au bout du fil. Bien qu'un petit peu étrange, l'échange téléphonique entre IA et humain se déroule naturellement et le rendez-vous est fixé. La personne du salon de coiffure n'a probablement aucune idée qu'elle parle à une IA. Ce qui est impressionnant c'est que l'assistant virtuel arrive à s'adapter en fonction des réponses de son interlocuteur humain.

Le futur proche de cette nouvelle fonctionnalité est clair. Nous ne prendrons même plus la peine de téléphoner pour prendre rendez-vous, ou pour commander un article. On demandera juste à notre assistant virtuel de le faire en notre nom. On pourrait même arriver à un stade où mon assistant virtuel appelle l'assistant virtuel du salon de coiffure pour prendre rendez-vous. L'humain est enlevé de l'équation.

Ainsi, notre assistant virtuel pourrait devenir extrêmement efficace à gérer des tâches de notre quotidien. Prendre rendez pour nous comme on l'a vue, mais également gérer notre calendrier, commander les produits qui nous manque en fonction de nos habitudes alimentaires, nous donner des infos en temps réels sur notre santé grâce à des capteurs biométriques. Nos lieux de vie seront également connectés à notre assistant si bien qu'il pourra gérer le chauffage, la lumière et tous les appareils faisant partie de l'internet des objets.

Notre IA personnelle nous connaitra de mieux en mieux à mesure qu'on

l'utilisera.

Tout comme les algorithmes d'Amazon, Google, Netflix ou encore Spotify apprennent basés sur nos préférences. Ce qui pourra amener à un scénario où avec suffisamment de données biométriques et suffisamment de puissance de calcul pour traiter ces données, un algorithme externe peut me connaître mieux que je ne me connais moi même. Il peut comprendre mes désirs, mes émotions, mes pensées, mes décisions. Il peut donc me proposer des choses qui seront parfaitement adaptées à mes goûts sans même que je ne le sache. Il peut également me contrôler et me manipuler bien plus facilement. Non seulement ses algorithmes sauront mieux ce que vous ressentez que vous même, mais ils sauront également pourquoi vous ressentez telle ou telle émotions et pourront donc prendre de meilleures décisions pour vous. Surtout lorsque vous n'avez pas envie de vous embêter. Par exemple si vous êtes en train de travailler sur un dossier important et votre conjoint vous appelle pour vous demander quel parfum de glace elle doit prendre au supermarché. Vous êtes concentrés et le parfum de la glace est la dernière chose dans la liste des priorités. Donc vous laissez votre assistant personnel prendre cette décision pour vous. Il saura exactement le parfum qu'il vous faut, basé sur les 150 glaces précédentes que vous avez mangé dans 107

l'année. 99% de nos décisions - y compris les choix de vie les plus importants comme qui épouser, quel travail choisir et quelle maison acheter - sont faites par les algorithmes complexes que nous appelons les sensations, les émotions et les désirs.

Donc en théorie, si des algorithmes externes arrivent à comprendre nos algorithmes internes, ils pourront suggérer 99% de nos décisions. Ce qui fait beaucoup !

Regardons le cas d'une des plus importantes décisions que l'on peut prendre dans sa vie. Qui épouser ? Dans la plupart des pays développés, plus d'un mariage sur deux se termine en divorce, avec des taux allant même jusqu'à 70% de divorce en Belgique. Pas génial. Voilà comment mon assistant personnel pourra répondre à la question. Je sors mon smartphone et je lui demande "*Qui épouser ?* ". Il me répond -

“*Je te connais depuis le moment où tu es né. J'ai lu tous les e-mails que t’as écrits.*

*J'ai écouté tous tes appels téléphoniques. Je me souviens de toutes tes relations amoureuses. Je peux te montrer les graphiques de ta tension artérielle pour chaque partenaire et rencontre sexuelle dans ta vie. Je connais aussi tes partenaires potentiels, comme je te connais. Je connais tes fantasmes, ce qui t'excite, ta personnalité, tes forces et faiblesses sentimentales. Sur la base de toutes ces informations, et sur les données de millions de relations réussies et infructueuses, je peux te recommander, avec une probabilité de 86%, que Samantha sera ta femme parfaite au lieu de Cassandre. Je te connais si bien que je sais que tu es déçu, car tu préfères Cassandre. Tu fais cette erreur, car tu donnes trop d'importance à l'apparence physique. La beauté est importante, mais tu lui en donnes trop. La beauté compte pour 9% du succès d’une relation, mais tes anciens algorithmes biochimiques, à cause de ce qui s'est passé dans la savane africaine où la beauté était synonyme de santé, donnent à l’importance de la beauté, un facteur de 27%, ce qui est beaucoup trop élevé à notre époque. Donc, même si tu penses Cassandre, choisis Samantha et tu seras heureux dans une relation amoureuse à la combinaison parfaite. ”*

Et il n’y a pas besoin que les décisions des algorithmes soient parfaites. En faite, il faut juste qu’ils soient un peu meilleurs que la moyenne des décisions humaines. Si on se rend compte que les recommandations des IA personnelles sont deux ou trois fois meilleures que les décisions que l’on prend par nous même, alors on les utilisera. Ça n’arrivera pas du jour au lendemain. Ce sera progressif. On demandera de plus en plus les avis de Google, Facebook, Amazon. On déléguera de plus en plus de décision. Et si on voit que leurs avis débouchent sur de meilleures décisions pour nous, alors on aura tendance à leur faire de plus en plus confiance.

Il sera également naturel d'éprouver de la sympathie pour nos assistants personnels. Le cerveau des humains est câblé pour ressentir de l’empathie pour des raisons évolutionnaires. Ce trait particulier n’est pas réservé à d’autres êtres humains puisque nombreux sont ceux qui élargissent le cercle de l’empathie aux chiens, chats 108

et autres animaux. D’une manière générale, on constate que plus l’animal a la

capacité de nous renvoyer des expressions, et donc des émotions, plus nous pouvons engendrer de l'empathie. C'est pour cette raison que la plupart des êtres humains ont beaucoup d'affection pour les chats, et très peu pour les carpes. Eh oui, regarder un chat dans les yeux n'a pas le même effet que croiser le regard d'une truite. Encore moins pour une grenouille ou une fourmi. Mais est-il possible de posséder de l'empathie pour un objet ou une entité artificielle ?

Grâce à des années de recherche, on sait aujourd'hui que la réponse est oui. Il existe d'ailleurs un terme pour ça : l'effet tamagotchi. Qui renvoi à ces petits animaux de compagnies virtuelles dans les années 90, en provenance tout droit du Japon. On a aussi des études montrant que les personnes âgées s'attachent à des compagnons robotiques dans les maisons de retraite. Et plus récemment, on sait que les gens s'attachent aux assistants virtuels : Siri, Alexa, Google Home. Et il y a également eu des cas où des soldats enterrent leur robot démineur lorsqu'ils sont trop endommagés, car ils ont travaillé à leur côté lors de nombreuse mission. Ils le perçoivent un peu comme un chien démineur.

Mais la preuve la plus éloquente d'une connexion entre humain et virtuelle se trouve au Japon avec "Loveplus". Un nombre grandissant de jeunes hommes japonais sortent avec des filles virtuelles. Bien qu'elles ne puissent interagir avec leur partenaire qu'en utilisant un script pré-écrit, ces filles virtuelles - Rinko, Manaka ou Nene - offrent une sorte de connexion émotionnelle instantanée. Elles peuvent embrasser, "tenir" la main de l'utilisateur, échanger des textos et même se mettre en colère si l'utilisateur ne répond pas à une conversation. Ce jeu est disponible sur les consoles portables nintendo et sur l'iPhone. Le Japon traverse une période de baisse démographique et un certain désintérêt de plus en plus important chez les jeunes générations vis-à-vis des relations de couple. Il n'est donc pas étonnant de voir le succès d'un jeu proposant à l'utilisateur d'avoir une copine virtuelle.

Nous allons donc potentiellement avoir des sentiments voire même tomber amoureux de nos assistants virtuels. Comme l'imagine le scénario du film "Her" de Spike Jonze sorti en 2013. Ou alors ils seront des sortes de "meilleurs potes".

Le rôle des assistants virtuels sera également très important dans l'éducation. Dans la mesure où notre assistant connaîtra beaucoup d'informations sur nos préférences, notre façon d'apprendre, nos facilités et nos défauts, il pourrait devenir un professeur personnalisé d'une efficacité bien plus grande que le système éducatif traditionnel. Ou même les cours en lignes. Le Google de 2030 ne sera pas un système pour trouver de l'information, mais un système dédié directement à faire apprendre à l'utilisateur.

109

**Armes autonomes**

Lorsqu'on mélange intelligence artificielle et guerre, on a tout de suite en tête les scénarios les plus dystopiques et apocalyptique que la Science Fiction nous ai offert comme Terminator et Matrix.

Un système d'armes létales autonome (SALA), ou robot tueur, est une machine capable d'avoir une action létale de manière automatisée sans intervention humaine.

Autrement dit, la décision de tuer a pour origine le système virtuel. Il peut s'agir d'un drone ou une autre machine. Mais cette définition n'est pas suffisamment précise pour faire l'unanimité. Des philosophes et chercheurs comme Peter Asaro et Mark Gubrud estiment que tout système d'armement capable d'exercer une force meurtrière sans opération, décision ou confirmation d'un superviseur humain peut être considéré comme autonome. Selon Gubrud, un système fonctionnant partiellement ou totalement sans intervention humaine est considéré comme autonome. Il souligne qu'un système armé n'a pas besoin de pouvoir prendre des décisions complètement par lui-même pour être appelé autonomes. Au lieu de cela, il devrait être considéré comme autonome tant qu'il est impliqué activement dans une ou plusieurs parties du processus menant à l'exécution. De la recherche de la cible au tir final.

Cependant, d'autres organisations fixent plus haut la norme du système d'arme autonome. Le ministère de la Défense du Royaume-Uni les définit comme des systèmes capables de comprendre des intentions et des directives complexes. Grâce à cette compréhension et à sa perception de

son environnement, un tel système est en mesure de prendre les mesures appropriées pour obtenir le résultat souhaité. Il est capable de décider d'un plan d'action, parmi un certain nombre de choix, sans dépendre de la surveillance et du contrôle humains.

En conséquence, la rédaction d'un traité entre États nécessite une définition communément acceptée de ce qui constitue exactement une arme autonome. Bien que de plus en plus de chercheurs, philosophes et organisations mettent garde face aux dangers d'une course à l'armement pour les armes autonomes.

Certaines personnes postulent que les armes nucléaires sont une des raisons majeures qui a permis la fin des grands conflits entre pays. Car ceux qui les possèdent savent les conséquences apocalyptiques qui découlent de leurs utilisations. Du coup, on pourrait se dire que si on construit des armes encore plus destructrices, basées sur des hordes de robots tueurs, alors on entrera dans une ère de paix sans précédent et l'idée même de guerre sera révolue.

Mais c'est un argument un petit peu léger … malheureusement, les guerres ne vont pas disparaître simplement, car nos armes sont trop destructrices. Et même si les conséquences d'une guerre à grande échelle dissuadent, comment s'assurer qu'un 110

chef d'État ne devienne pas fou et lance une attaque dévastatrice sur un pays ennemi ?

Donc si les conflits sont inévitables, la question c'est de savoir comment minimiser la souffrance infligée aux humains ? Et l'intelligence artificielle est une réponse possible. Si les guerres consistent simplement en des machines combattant d'autres machines, alors aucun soldat humain ou civil n'a besoin d'être tué (en théorie).

De plus, les futurs drones et autres systèmes d'armes autonomes peuvent être en principe bien plus rationnel que des soldats rongés par des émotions conflictuelles.

Équipés de capteurs surhumains et sans peur de se faire tuer, ils

pourraient rester calmes même dans le feu de l'action, et seront moins susceptibles de tuer accidentellement des civils. Sur le papier, cela sonne comme une bonne chose à mettre au point. Seulement, lorsqu'on parle de système automatisé, on fait face à de potentiels bugs qui, associés à de l'armement léthal, peuvent avoir de très lourdes conséquences.

C'est déjà arrivé par le passé, et plus d'une fois. Voici un exemple probant tiré du livre "Life 3.0" de Max Tegmark. Le vol 655 d'Iran Air était un vol commercial assurant la liaison entre Téhéran et Dubaï. L'Airbus est abattu le 3 juillet 1988 au-dessus du golfe Persique par un tir de missiles provenant du croiseur américain USS

Vincennes, qui disposait d'un système autonome de détection et d'autodéfense. La catastrophe, qui fit 290 victimes civiles, dont 66 enfants, serait due à une erreur du système autonome qui a confondu l'Airbus avec un avion de combat F-14. Toutefois, il faut noter que le système autonome n'a pas pris lui même la décision de tirer. Le tir fut ordonné par le commandant William Rogers. La décision finale a été prise par un humain, qui a accordé trop de confiance en ce que lui disait le système autonome.

L'USS Vincennes, à l'origine de la catastrophe du vol 655 Iran Air le 3 juillet 1988.

Les systèmes d’armes autonomes existent depuis plusieurs décennies, mais uniquement à usage défensif. Ils peuvent identifier et attaquer de manière autonome les missiles, les roquettes, les tirs d'artillerie, les aéronefs et les navires de surface selon des critères définis par l'opérateur humain. Des systèmes similaires existent pour les chars russe, israélien et allemand. Plusieurs types de tourelles pouvant tirer sur des humains et des véhicules sont utilisés en Corée du Sud et en Israël aux frontières. De nombreux systèmes de défense antimissile, tels que Iron Dome, disposent également de capacités de ciblage autonomes. La principale raison de ne pas avoir une décision humaine dans ces systèmes est la nécessité d'une réponse rapide en cas d’attaque.

En revanche il n’existe aucun système autonome d’attaque sans intervention humaine, mais la technologie existe déjà et de nombreux projets militaires sont en phase de prototype. Les véhicules aériens de combat sans pilote font partie des systèmes dotés d’une plus grande

autonomie, par exemple : exécuter une cible quand il est autorisé par le commandement de la mission. Il peut également se défendre contre les avions ennemis. L'avion militaire Northrop Grumman X-47B peut décoller et atterrir sur des porte-avions sans assistance humaine. Selon le journal "The Economist", à mesure que la technologie progressera, les applications futures des véhicules sous-marins autonomes pourraient inclure le déminage, la pose de mines, la mise en réseau de capteurs anti-sous-marins, et le réapprovisionnement de sous-112

marins habités. La Russie développe une torpille nucléaire autonome intercontinentale nommée "Poseidon" capable entre autres d'exploser à proximité du littoral d'un pays ennemi, générant ainsi un tsunami d'eaux radioactives sur une ville. Pas le genre d'arme que l'on souhaite voir à l'action.

Le Northrop Grumman X-47B en opération autonome au-dessus de l'Atlantique.

Est-ce une bonne idée de donner la décision finale d'éliminer une cible à une machine ? Ou devons-nous toujours faire en sorte qu'un humain

donne l’ordre ?

Le 27 octobre 1962, lors de la crise des missiles à Cuba, le sous-marin soviétique B-59 se trouvait près de Cuba, dans les eaux internationales. L’équipage n’avait plus aucun contact avec Moscou depuis plusieurs jours et ne ils savaient pas si la troisième guerre mondiale avait déjà commencé. Puis les Américains ont commencé à lâcher des charges légères, simplement destinés à forcer le sous-marin à faire surface et à partir. Mais ils ignoraient que le B-59 possédait une torpille nucléaire autorisée à être lancée sans l’accord de Moscou. Et le capitaine Savitski décida de la lancer. Vous devez surement vous demander pourquoi nous n’avons pas connu la 3e guerre mondiale ? Tout simplement, car la décision devait être approuvée par les trois officiers à bord. Deux d’entre eux ont voté pour. Le troisième a refusé. Peu d’entre nous ont entendu parler de Vasili Arkhipov, malgré le fait que sa décision ait évité la troisième guerre mondiale. Mais du coup, si jamais le sous-marin B-59 avait possédé un système de lancement autonome, la fin du monde aurait peut-être eu lieu le 27

octobre 1962.

Photo de Vasili Arkhipov, l'officier ayant empêché une probable 3e guerre

mondiale. ©National Geographic CC-BY-SA-4.0

Ce n’est pas le seul incident ayant failli déboucher sur la troisième guerre mondiale. Le 9 septembre 1983, un système d'alerte soviétique automatisé signala que les États-Unis avaient lancé cinq missiles nucléaires en direction de l’Union soviétique, laissant à l'officier Stanislav Petrov simplement quelques minutes pour décider si c'était une fausse alerte. Il arriva à la conclusion que si les États unis attaquaient l’URSS, il ne le ferait pas avec seulement cinq missiles nucléaires. Il décida donc de signaler à ses supérieurs que c'était une fausse alerte. En effet, il s'avéra que le système de détection prit la réflexion du soleil sur les nuages pour des missiles. Encore une fois, si un système d’attaque autonome avait été en place à ce moment-là, il aurait certainement déclenché une pluie de missile nucléaire sur les

États unis.

On peut voir que dans certaines circonstances, un système d'armement autonome permettrait d'éviter des dommages collatéraux et des pertes civiles. Dans le cas de la lutte contre le terrorisme, si le but d'une mission est d'éliminer un des leaders d'une organisation terroriste, un système autonome pourrait localiser, traquer et éliminer uniquement la cible, sans faire sauter un immeuble plein de civiles. Comme on l'a vu 114

dans les parties précédentes, l'intelligence artificielle a fait d'énormes progrès dans la reconnaissance d'image. Elle est capable d'identifier un individu sur des images de faible qualité, en mouvement, basé sur le visage, la posture et d'autres paramètres.

Dans beaucoup de cas de figure, une intelligence artificielle a un niveau super-humain pour identifier une cible, ce qui dans le cas des armes autonomes, baisserait considérablement le taux d'erreur, car elle pourrait se rapprocher suffisamment près de la cible pour l'éliminer, sans faire d'autres victimes.

Par contre, le manque "d'humanité" derrière l'acte de tuer, peut entraîner des conséquences désastreuses. Certains massacres ethniques dans l'histoire ont fait des milliers de victimes, rasant des villages entiers. Mais certains soldats, face à l'ordre de tuer, par exemple, tous les musulmans d'un secteur, comme c'était le cas durant la guerre du Kosovo, ont ressenti de l'empathie pour les femmes et les enfants. Si l'ordre avait été donné à des armes autonomes, le pourcentage de décès ethnique aurait été virtuellement de 100% tant ces systèmes sont efficaces. Aucun remords. Aucune empathie. Aucune hésitation. Si la machine voit un individu qui correspond à ce qu'elle a appris en machine learning de ce qu'est telle ou telle ethnie, elle fera feu. On n'ose imaginer ce qu'Hitler aurait pu faire à la population juive s'il avait eu accès à des armes autonomes.

Si une puissance militaire majeure développe des intelligences artificielles au service de système autonome, une course à l'armement mondiale est pratiquement inévitable. Car même si les États-Unis et l'Europe annonce un traité stipulant l'interdiction de produire des armes

autonomes, est-ce que la Chine ou la Russie feront de même ? Sachant très bien cela, les États unis ne prendront pas le risque d'être en retard militairement sur de potentiels futurs ennemis, donc juste au cas où, ils continueront à développer des armes autonomes.

115

Le MQ-1 Predator de l'armée Américaine ne prend encore pas la décision de tuer.

Mais jusqu'à quand ?

Le point final de cette trajectoire technologique est évident : les armes autonomes deviendront les Kalachnikovs de demain. Contrairement aux armes nucléaires, elles ne nécessitent pas de matières premières coûteuses ou difficiles à obtenir, donc elles deviendront omniprésentes et bon marché. Ce ne sera qu'une question de temps jusqu'à ce qu'elles apparaissent sur le marché noir et donc dans les mains de terroristes, des dictateurs souhaitant mieux contrôler leur population, des seigneurs de

guerre souhaitant un nettoyage ethnique, etc. Les armes autonomes sont idéales pour des tâches telles que les assassinats, la déstabilisation d'une nation, et tuer de manière sélective un groupe ethnique particulier. On peut également souligner l'inquiétude vis-

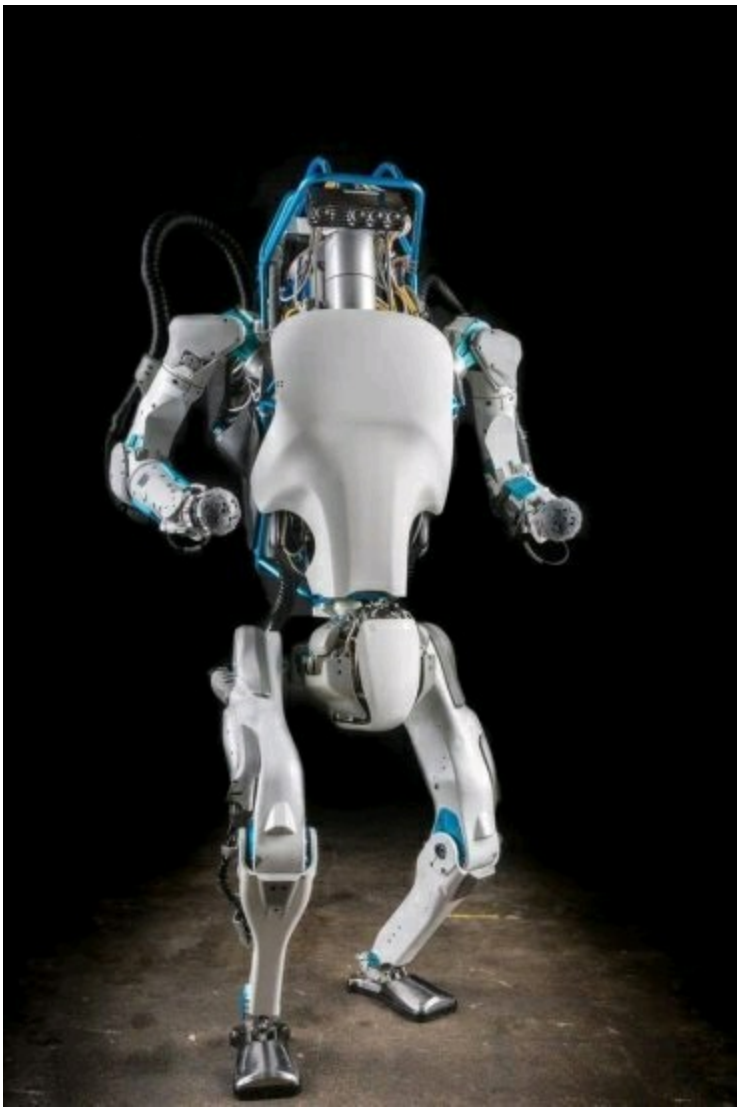
à-vis de possible piratage de système d'arme autonome. Des terroristes pourraient pirater un robot tueur d'une armée pour que ce dernier obéisse à leurs ordres. En réalité, il est déjà possible aujourd'hui de concevoir un drone tueur. Il suffit d'un drone lambda que l'on trouve sur Amazon, d'un logiciel de reconnaissance faciale, et d'une petite charge explosive. Pour très peu d'argent, et sans avoir besoin de nouvelle technologie, il est possible d'ordonner à un drone de tuer une cible spécifique. Voire même de fabriquer un essaim de drones tueur pour commettre un massacre, ou éliminer de nombreuses cibles.

Ce n'est pas vraiment un scénario à la Terminator qu'il faut craindre, mais plutôt celui d'un groupe terroriste qui programme un robot tueur afin de commettre un 116

massacre dans un stade de foot. Et il est bien plus difficile de stopper une machine qu'un humain. Ou encore un suprémaciste blanc néonazi membre du Klu Klux Klan qui ordonne à un drone autonome armé de tuer toutes les personnes de couleurs qu'il détecte.

Les drones actuels pourraient être armés dans le but de faire des victimes.

Et lorsque l'on voit les progrès en robotique au cours de ces dernières décennies, notamment par l'entreprise Boston dynamics, on peut être inquiet. Il y a dix ans, la tâche la plus compliquée pour les ingénieurs en robotique était de faire marcher leur robot sur deux jambes. Aujourd'hui, le robot Atlas peut courir, monter des escaliers, rester debout malgré un trébuchement et faire des saltos arrières. Et les robots quadrupèdes comme Spot sont également très impressionnant. Si vous avez vu l'épisode 5 de la saison 4 de Black mirror, vous comprenez que ce n'est pas une bonne idée de permettre à de tel robot le droit de tuer sans ordre humain.



Atlas est un type de robot bipède de Boston Dynamics ayant fait des progrès très impressionnants. CC BY-SA 4.0

Les armes autonomes pourraient engendrer finalement plus de conflits que par le passé, car elles baissent le seuil de décision d'entrer en guerre ou non. Étant donné qu'il y a moins de soldats humains en jeu, les gouvernements pourraient être plus enclins à régler des conflits par la guerre.

Les armes autonomes sont au coeur d'un débat aux quatre coins du monde, en particulier sur le risque de "robots tueurs". Le groupe "Campaign to Stop Killer Robots" s'est formé en 2013. En juillet 2015, plus de 1 000 experts en intelligence artificielle ont signé une

lettre mettant en garde la menace d'une course aux armements en matière d'intelligence artificielle militaire et appellent à l'interdiction des armes autonomes. La lettre a été présentée à Buenos Aires lors de la 24e Conférence internationale sur l'intelligence artificielle, signée par Stephen Hawking, Elon Musk, Steve Wozniak, Noam Chomsky, et le fondateur de Google DeepMind Demis Hassabis, entre autres.

118

Tout comme la plupart des chimistes et biologistes n'ont aucun intérêt à fabriquer des armes chimiques ou biologiques, la plupart des chercheurs en intelligence artificielle n'ont aucun intérêt à construire des armes autonomes et ne veulent pas que de tels projets ternissent l'image de l'intelligence artificielle qui pourrait créer une réaction publique majeure contre l'IA, ce qui ralentirait les futurs avantages. Les chimistes et biologistes ont largement soutenu les accords internationaux qui ont interdit avec succès les armes chimiques et biologiques, tout comme la plupart physiciens ont soutenu les traités interdisant les armes nucléaires dans l'espace. On peut donc espérer que les chercheurs en intelligence artificielle limitent le développement des armes autonomes et que des réglementations internationales verront le jour.

Le 12 septembre 2018, le Parlement européen a adopté une résolution appelant à une interdiction internationale des systèmes d'armes autonomes létales. La résolution a été adoptée par 82% des membres qui ont voté en sa faveur.

119

3.

Demain : Intelligence artificielle générale

120

3. Demain : Intelligence artificielle générale

Pourquoi est-ce si dur ?

Il existe dans la nature différents types d'organisme qui semble être très doué à des tâches spécifiques. En d'autres termes, ils ont une intelligence spécialisée. Les abeilles semblent compétentes à faire des ruches. Les castors semblent compétents pour faire des barrages. Mais si un castor observe des abeilles fabriquer une ruche, il ne va pas essayer de faire pareil. Il semble donc qu'il y a un manque de transfert de compétence. Par opposition, nous, humains, pouvons observer les castors en train de construire un barrage, et répliquer à notre façon cette compétence. Nous pouvons observer comment les abeilles construisent leurs ruches et faire de même. Nous avons une intelligence générale. Et en regardant l'ampleur de notre savoir-faire, il semblerait que nous ayons l'intelligence générale la plus vaste du règne animal. C'est ce type d'intelligence que certains chercheurs tentent de construire dans leur atelier.

Pour faire simple, si une intelligence artificielle est capable de développer un système d'apprentissage qui lui permet d'apprendre par elle même des dizaines et des dizaines de choses, elle sera considérée comme ayant une intelligence générale (ou forte). Par opposition à l'intelligence artificielle limitée (ou faible) que l'on peut voir aujourd'hui autour de nous. Elle aura la capacité d'appliquer sa technique d'apprentissage à un nombre virtuellement illimité de domaines, et pourra atteindre le niveau d'intelligence d'un humain. Cette capacité d'apprendre à apprendre est appelée meta-apprentissage.

C'est intéressant de noter que ce qui est souvent intuitivement considéré comme le plus difficile pour nous humain se révèle être super simple pour une intelligence artificielle. Multiplier des nombres de dix chiffres en une fraction de seconde c'est réservé aux autistes spéciaux, mais c'est hyper facile pour un ordinateur. Faire la différence entre un chien et un chat c'est un jeu d'enfant pour 99,9% des êtres humains, mais pour une intelligence artificielle c'est d'une complexité insurmontable.

Faire de l'IA capable de battre n'importe quel humain aux échecs ? C'est fait. Faites-en une capable de lire un paragraphe d'un livre d'un enfant de six ans et ne pas simplement reconnaître les mots, mais en comprendre la signification ? Google dépense actuellement des milliards de dollars pour essayer de le faire. Les tâches difficiles pour nous comme le calcul,

l'analyse du marché économique et la traduction linguistique en 150 langues sont extrêmement faciles à comprendre pour un ordinateur, tandis que les choses simples comme la vision, la compréhension du 121

langage naturel et la perception sont incroyablement difficiles. Ou, comme le dit l'informaticien Donald Knuth : "*Jusqu'à présent, l'intelligence artificielle arrive à faire quasiment tout ce qui requiert de la penser, mais échoue à faire ce que les humains et animaux font "sans penser*"".

Ces choses qui nous semblent si faciles sont en réalité incroyablement compliquées. Elles semblent faciles parce que ces compétences ont été optimisées par des centaines de millions d'années d'évolution à travers la sélection naturelle. Lorsque vous tendez la main vers un objet, les muscles, les tendons et les os de votre épaule, de votre coude et de votre poignet effectuent instantanément une longue série d'opérations physiques, en conjonction avec vos yeux, pour vous permettre de bouger votre main sur trois dimensions. Cela semble sans effort pour vous parce que le logiciel dans votre cerveau a eu des millions d'années d'entraînement à travers des milliers de générations. Par contre, multiplier des grands nombres et jouer aux échecs sont des activités relativement récentes pour le cerveau humain, il n'a pas eu des millions d'années pour se perfectionner donc il est inefficace.

C'est pour cette raison que l'apprentissage est un élément clé dans la création d'une intelligence artificielle. Si on pouvait mettre une IA dans une boite avec tout un tas de données sur le monde plus une dose de machine learning, pendant cent millions d'années, elle ressortirait surement avec un niveau d'intelligence dépassant de loin celui d'un humain. Mais ce serait quand un peu long et la plupart des chercheurs travaillant sur le développement d'une IA générale espèrent le faire bien plus rapidement.

Donc l'élément clé pour la création d'une intelligence artificielle générale c'est l'apprentissage. Comme on l'a vu dans la partie précédente, le machine learning et le deep learning sont des techniques efficaces pour faire apprendre à des algorithmes certaines fonctionnalités. Mais des

tonnes de données sont nécessaires pour nourrir ses algorithmes, ce qui est très différent de la façon dont un être humain est capable d'apprendre. Pour nous, un ou deux exemples et nous sommes capables de tirer des conclusions rapidement. Par exemple, pour reconnaître de la neige sur une photo.

Pour un enfant de 10 ans vivant sur une île tropicale qui n'a jamais vu de la neige, il lui suffira de voir une ou deux photos et il aura appris. Une IA doit analyser des milliers de photos pour obtenir un résultat proche des 100% lorsqu'on lui demande de reconnaître un paysage enneigé. Une intelligence artificielle générale devrait donc être capable d'apprendre une tâche avec très peu d'exemples à disposition. Et c'est quelque chose qui semble être très compliqué aujourd'hui.

Un autre problème concernant l'apprentissage c'est bien entendu la généralisation des connaissances qu'un système artificiel est capable d'apprendre. D'où le terme 122

"intelligence artificielle générale" et également le transfert d'apprentissage. C'est-à-

dire que lorsqu'un enfant commence à devenir bon au jeu vidéo "Mario", s'il commence à jouer à Sonic, il aura déjà des acquis techniques qui seront applicables sur tous les jeux de plateforme similaire. Si l'on prend une intelligence artificielle qui a impressionné par ces capacités, disons Alpha Zero de Google deepmind. Il sera capable de battre n'importe quel humain au jeu de Go jusqu'à la fin des temps, mais il ne sait pas commander une pizza en ligne, ou conduire une voiture. Même s'il a néanmoins démontré des capacités de transfert d'apprentissage en maîtrisant rapidement les échecs en plus du jeu de Go, mais ces capacités restent limitées.

Depuis le lancement de la recherche sur l'intelligence artificielle en 1956, la recherche dans la création d'IA générale s'est ralentie au fil du temps. Une des explications est que les ordinateurs ne disposent pas de suffisamment de mémoire ou de puissance de traitement pour simuler la complexité du cerveau humain, et donc développer une intelligence comparable.

D'autres raisons possibles ont été avancées pour justifier les difficultés rencontrées comme la nécessité de bien comprendre le cerveau humain par le biais de la psychologie et de la neurophysiologie qui ont empêché de nombreux chercheurs d'imiter la structure du cerveau humain dans un cadre artificiel. Il y a également le débat sur l'idée de créer des machines avec des émotions. Il n'y a pas d'émotions dans les modèles actuels et certains chercheurs affirment que la programmation d'émotions dans des machines leur permettrait d'avoir leur propre esprit. L'émotion résume les expériences des humains, car elle leur permet de se souvenir de leurs expériences. Le chercheur David Gelernter a écrit qu'aucun ordinateur ne sera créatif s'il ne peut simuler toutes les nuances de l'émotion humaine.

Mais l'une des difficultés qui occupent grandement les chercheurs c'est la question du "bon sens". C'est-à-dire comment s'assurer qu'une intelligence artificielle puisse comprendre le monde et faire preuve de bon sens dans sa résolution des problèmes qu'elle rencontre ? Pour les humains, le bon sens est un ensemble de faits que tout le monde connaît et prend pour acquis lorsque l'on exécute une tâche. Par exemple, "*un enfant restera plus jeune que ses parents jusqu'à la fin de sa vie*", "*est-ce possible de faire une salade avec du polyester ?*", "*Il est impossible que mon grand-père n'ait jamais eu d'enfant*" etc.

Ce sont des questions absurdes, que n'importe quel humain de plus de 2 ans est capable de savoir. Mais une intelligence artificielle n'en a aucune idée et si elle n'apprend pas chaque réponse, elle ne pourra pas les déduire d'elle même. Par exemple, si je demande à une intelligence artificielle : *"Est-ce que Monsieur Martin est plus grand que son bébé ?"*, l'IA va chercher à résoudre cette question en trouvant 123

les données sur la taille des deux êtres humains. Elle me répondra : "*Monsieur Martin fait 1m80, son bébé fait 70 cm donc la réponse à votre question est que Monsieur Martin est plus grand*". Mais ce n'est pas la meilleure façon de répondre à cette question puisque n'importe qui sait qu'un adulte sera forcément plus grand que son bébé. Pas besoin de comparer leur taille pour savoir qui est plus grand.

Alors on pourrait se dire que la solution est simple. Il suffit d'apprendre à

l'IA que les adultes sont plus grands que leur bébé, dans tous les cas. Mais ce n'est qu'un seul fait de bon sens parmi des milliards d'autres. Enseigner tout le bon sens des humains à une IA est une tâche incommensurable. Et même si on y parvient, comment être sûr que l'on n'a rien oublié ? Sachant qu'un oubli pourrait avoir des conséquences désagréables pour ne pas dire tragique.

Imaginons un robot ménager dans votre maison ayant une intelligence artificielle très avancée, proche d'une IA générale. Vous lui demandez de préparer le dîner, mais le robot ne trouve aucun aliment dans le frigo. Soudain, votre chat passe sous ses yeux. 10 minutes plus tard, le robot vous sert un ragoût de chat. Ce n'est pas vraiment ce que vous attendiez, mais vu que le robot n'a pas le bon sens de savoir qu'on ne mange pas nos animaux de compagnie, il a fait ce qu'il pensait être la bonne action par rapport à votre requête.

Autre exemple avec une super intelligence capable de résoudre tous nos problèmes. Si on lui demande d'arrêter la faim dans le monde, elle pourrait très bien arriver à la conclusion que d'éliminer tous les êtres vivants sur la planète est le meilleur moyen d'empêcher un organisme de ressentir la sensation de faim. Car si plus d'organismes, plus de faim. Logique. Et si on lui dit que ce n'est pas ce qu'on lui demande, elle nous répondra que c'est exactement ce qu'on lui a demandé. Dans ce cas, il aurait fallu préciser "Élimine la faim dans le monde sans tuer tous les êtres vivants". Pour nous humain, c'est du bon sens. Évidemment que lorsque l'on souhaite éliminer la faim dans le monde, on souhaite également préserver la vie sur Terre. Mais pour une IA, aussi intelligente soit-elle, le bon sens n'existe pas.

Il nous faut donc trouver un moyen de concevoir une intelligence artificielle capable de déduire et de raisonner afin de tirer du bon sens les situations qu'elle rencontre. Et c'est un sujet de recherche de plus en plus répandu. Peu de temps avant son décès, le cofondateur de Microsoft Paul Allen, a injecté 125 millions de dollars dans son laboratoire de recherche dans le but premier de percer le mystère du bon sens chez une machine.

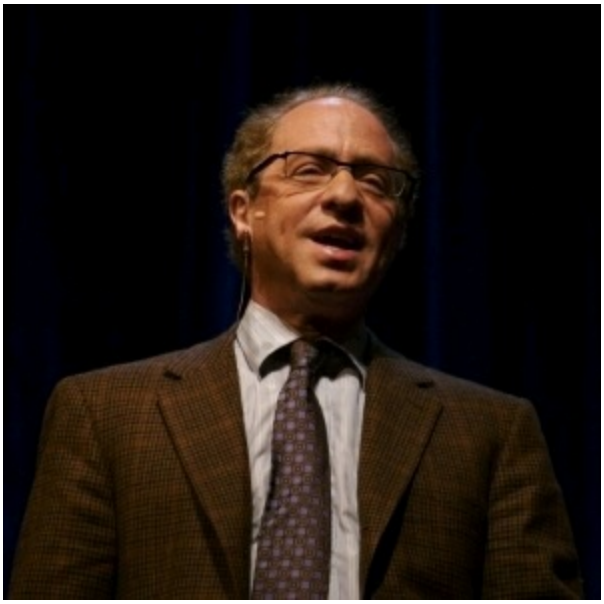
En résumé, il n'y a absolument aucune garantie que nous parvenions à

concevoir une intelligence artificielle ayant les compétences générales de l’intelligence humaine.

124

Mais il n’y a pas non plus d’argument prouvant que cela est impossible. Nous ne savons pas à quelle distance nous sommes de la ligne d'arrivée. L'architecture matérielle, les algorithmes et les techniques d’apprentissages ont permis des progrès fulgurants dans le domaine et il semblerait que chaque année, nous faisons plus de progrès que l’année précédente. Ce qui indique une croissance exponentielle. En d’autres termes, nous ne pouvons pas rejeter la possibilité que l’intelligence artificielle atteigne éventuellement le niveau humain et au-delà. Voyons donc les implications que cela pourrait entraîner.

125



3. Demain : Intelligence artificielle générale

Les moyens pour y arriver

Même si la tâche est extrêmement difficile et que certains chercheurs doutent de la faisabilité de créer une intelligence artificielle générale, d’autres travaillent ardemment sur des pistes qui semble prometteuses. La

recherche est extrêmement diversifiée et fait souvent figure de pionnière. Le chercheur Ben Goertzel pense que la durée approximative nécessaire pour créer une IA forte varie de 10 ans à plus d'un siècle, mais le consensus qui s'est dégagé au sein de la communauté de chercheurs semble être que la prédiction mise en avant par Ray Kurzweil dans son livre écrit en 2005 "La singularité est proche" est plausible. C'est-à-dire entre 2015 et 2045. Même si beaucoup de chercheurs en intelligence artificielle doutent que les progrès soient aussi rapides. Les organisations qui poursuivent explicitement la création d'une IA forte incluent le laboratoire suisse IDSIA, Nnaisense, Vicarious, la OpenCog Foundation, SingularityNet, Adaptive AI, LIDA, Numenta et le Redwood Neuroscience Institute associate. De plus, des organisations telles que le Machine Intelligence Research Institute et OpenAI ont été créées pour influencer le développement de l'IA générale. Enfin, des projets tels que the Human Brain project ont pour objectif de construire une simulation fonctionnelle du cerveau humain.

Le futurologue Ray Kurzweil est célèbre pour ses prédictions sur l'intelligence artificielle ©null0 CC BY-SA 2.0

Donc voici les pistes pour arriver un jour à la mise au point d'une intelligence 126

artificielle générale.

**Une puissance de calcul similaire à celle du cerveau**

Si l'on souhaite créer une intelligence artificielle générale, on peut d'abord opter pour une approche de comparaison. C'est-à-dire se tourner sur les exemples, dans la nature, qui ont produits des intelligences générales. On constate en faisant cela que le point commun de toutes les intelligences générales de la planète, les humaines bien sûr, mais aussi les autres primates, mammifères, etc., c'est qu'ils possèdent un cerveau. Cela peut paraître trivial, mais c'est une première étape nécessaire afin de construire un système artificiel capable de générer une intelligence générale.

Le cerveau est à ce jour, l'objet que nous connaissons le plus complexe dans l'univers. Mais les progrès en neuroscience grâce

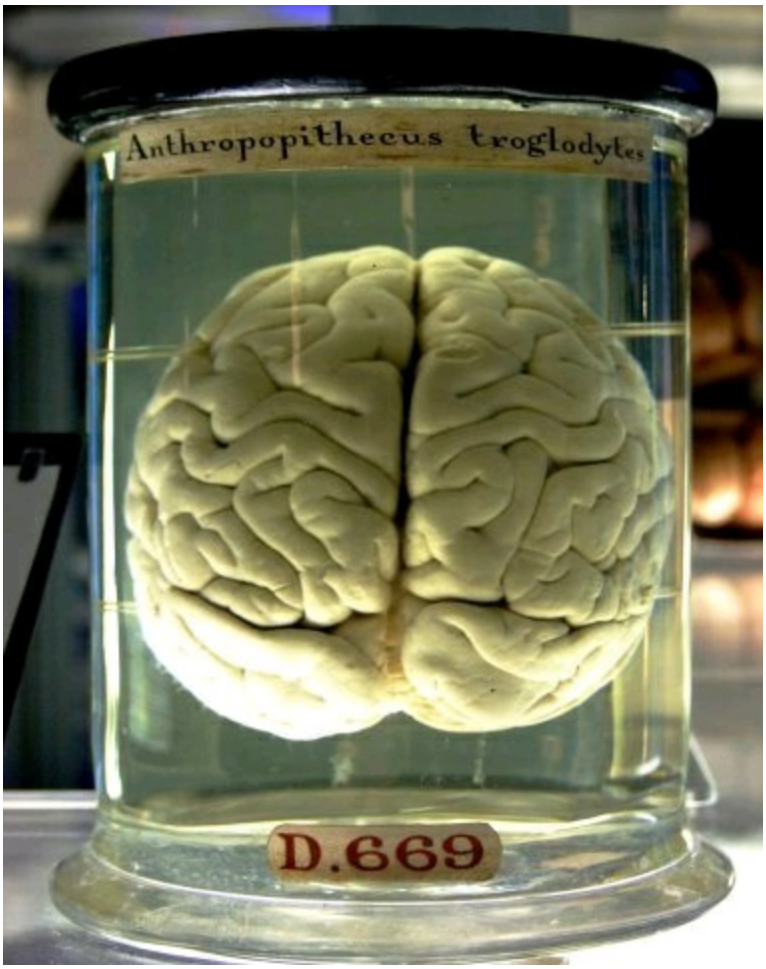
notamment à des outils de neuro-imagerie plus performante, nous ont permis de comprendre de nombreuses fonctionnalités du cerveau. Ceci étant dit, nous n'avons pas percé tous les mystères entourant cet organe si particulier.

La première chose que l'on peut se demander c'est qu'elle est la puissance de calcul du cerveau humain ? Afin de le comparer avec celle de nos machines. Il se trouve que notre cerveau possède un grand nombre de synapses. Chacun des 1011

(cent milliards) neurones possède en moyenne 10 000 connexions synaptiques avec d'autres neurones. Ce qui signifie 100 à 500 trillions de synapses et les signaux sont transmis le long de ces synapses à une fréquence moyenne d'environ 100 Hz. Une façon d'exprimer la puissance de calcul du cerveau est de mesurer le nombre total de calculs par seconde (cps) que le cerveau peut générer. Plusieurs chercheurs se sont penchés sur cette tâche et sont parvenus à un nombre entre 1016 et 1017 cps.

Une autre façon de calculer la capacité totale consiste à regarder une partie du cortex qui remplit une fonction que nous savons simuler sur des ordinateurs. Hans Moravec a effectué ce calcul à l'aide de données sur la rétine humaine en 1997. Il a obtenu la valeur 1014 cps pour le cerveau humain dans son ensemble. À titre de comparaison, si un "calcul" était équivalent à une "opération à virgule flottante" - une mesure utilisée pour évaluer les superordinateurs actuels - 1016 cps équivalents à 10

petaFLOPS.




Le cerveau humain est capable d'une puissance de calcul de 1016 cps.
©Gaetan Lee CC BY 2.0

Est-ce que nous sommes loin d’avoir des ordinateurs capables de produire 1016

cps ? Non, car aussi surprenant que cela puisse paraître, nous avons déjà des superordinateurs plus puissants. Avec une performance maximale de 200 pétaflops, soit 200 000 milliards de calculs par seconde, Summit d’IBM et le superordinateur le plus puissant du monde. Il utilise 4 608 serveurs de calcul contenant deux processeurs IBM Power9 à 22 cœurs et six unités graphiques Nvidia Tesla V100. Summit consomme 13 mégawatts, tandis que le cerveau n’a besoin que de 20 watts d'énergie pour fonctionner.

À l'heure de l'écriture de ces lignes (janvier 2019), Summit d'IBM est le

superordinateur le plus puissant au monde capable de 200 petaflops ©Carlos

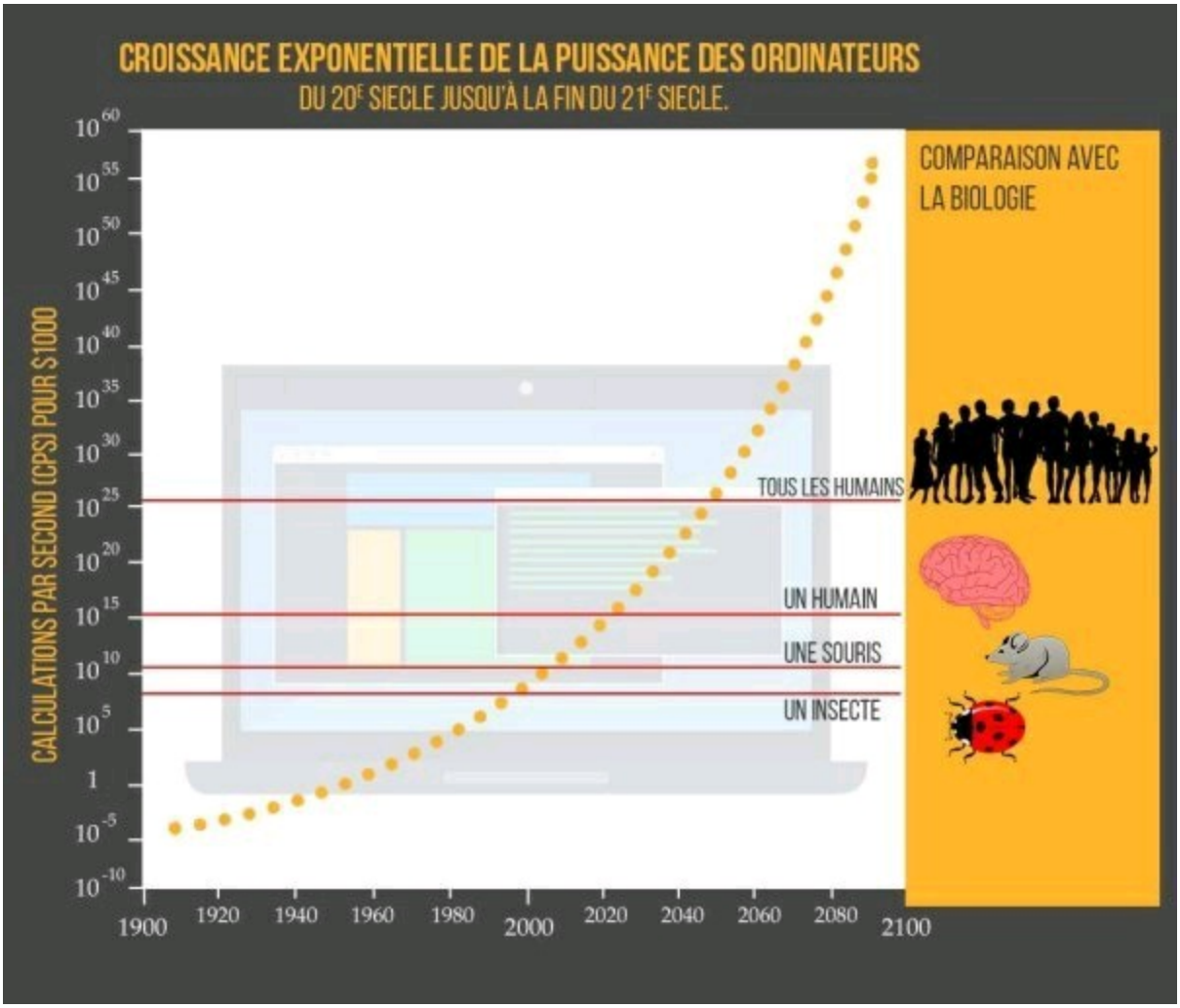
Jones/ORNL CC BY 2.0

Mais si nous avons déjà un super ordinateur qui a une puissance de calcul 20 fois supérieure à celle du cerveau humain, pourquoi n'avons-nous toujours pas d'intelligence artificielle générale ?

Alors il faut déjà prendre en compte que Summit a coûté 200 millions de dollars à construire, ce qui limite forcément son usage. Historiquement, les superordinateurs ont été utilisés pour des simulations météorologique, atmosphérique, moléculaires, astrophysiques ou encore pour gérer efficacement les réserves nucléaires. Ils ne sont pas accessibles à tous les chercheurs travaillant dans l'intelligence artificielle bien qu'il existe des projets visant à simuler la complexité du cerveau humain.

Ce qui signifie que tant que l'accès à la puissance du calcul du cerveau coûtera plusieurs centaines de millions, les progrès pour créer une IA générale seront extrêmement limités. Ray Kurzweil suggère que

nous devons plutôt regarder combien de cps nous pouvons acheter pour 1 000 $. Lorsque ce nombre atteindra le niveau de calcul du cerveau humain (1016) cela signifiera qu'une IA générale pourrait entrer en scène rapidement. En 2019, il est possible d'obtenir en moyenne 10 trillions de cps (1013) pour un ordinateur possédant un microprocesseur Intel i7 par exemple. Ce qui correspond à la puissance de calcul du cerveau d'une souris. Ray Kurzweil prédit depuis longtemps que nous arriverons à 1016 pour 1 000$ en 2025, ce qui semble cohérent avec la croissance exponentielle de la puissance informatique suivant la loi 129




de Moore.

Courbe de la croissance exponentielle de la puissance de calcul vu par Ray

Kurzweil dans "The Singularity is Near".

Enfin il ne faut pas croire qu'il suffit d'allumer un superordinateur de 200

pétaflops et le laisser tourner pendant 2 semaines pour voir émerger une intelligence artificielle générale. La puissance de calcul est une étape importante, mais la façon dont un ordinateur fonctionne détermine ce qu'il produit. En d'autres termes, nous avons déjà construit une machine ayant la puissance de calcul du cerveau, mais nous n'avons pas émulé comment il fonctionne.

**Simuler le cerveau humain**

Une approche populaire discutée pour créer une intelligence artificielle générale est l'émulation du cerveau humain. L'idée est de concevoir un modèle cérébral en cartographiant en détail un cerveau biologique et en copiant sa structure dans un système informatique. En théorie, l'ordinateur devrait exécuter ensuite une simulation si fidèle à l'original qu'il se comportera essentiellement de la même manière que le 130

cerveau d'origine, ce qui produira une intelligence générale comparable à un humain.

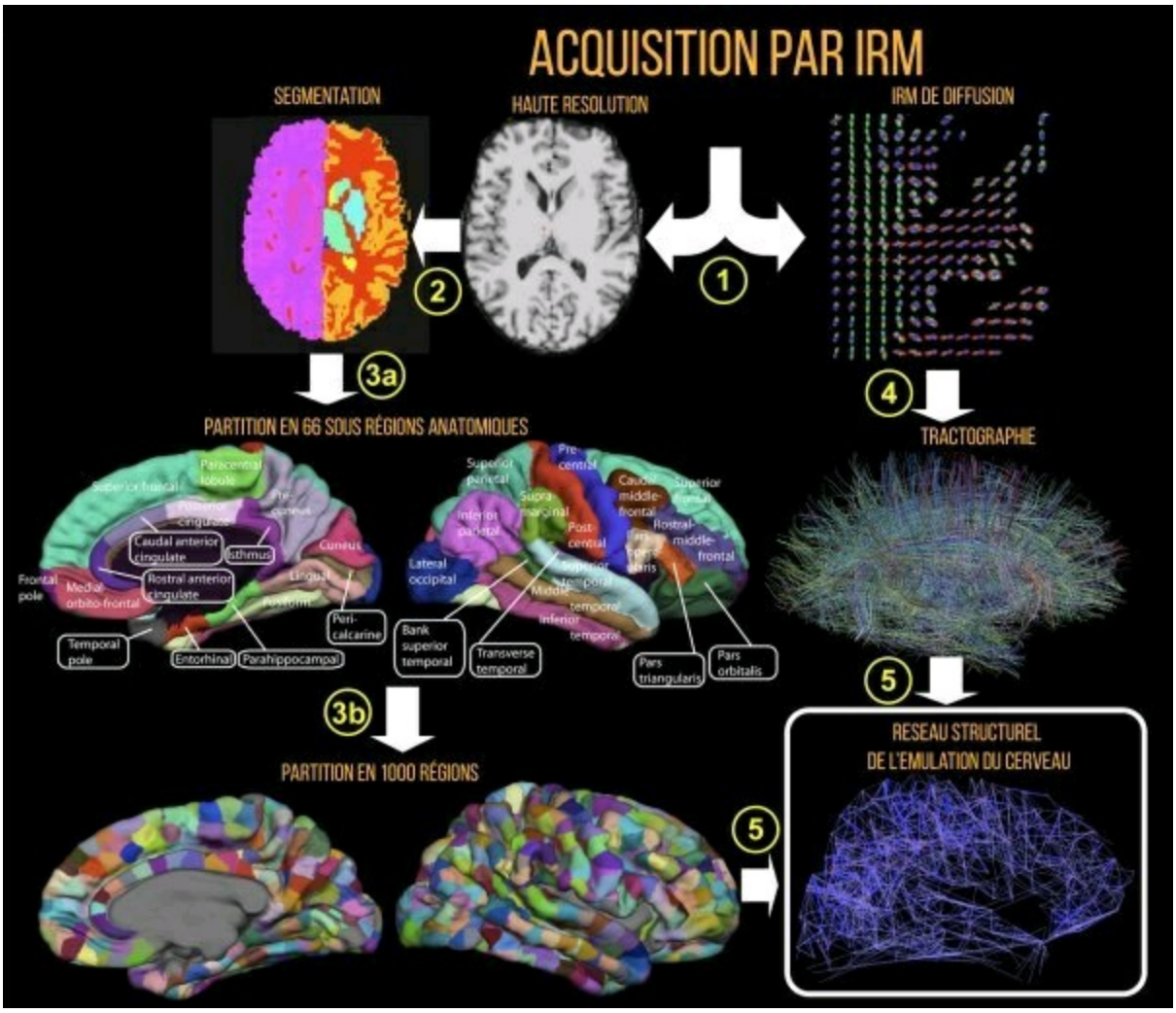
L'émulation du cerveau entier est un sujet d'étude à l'origine issu des neurosciences pour obtenir des simulations du cerveau à des fins de recherche médicale. Les technologies de neuro-imagerie susceptibles de fournir une compréhension détaillée nécessaire du cerveau s'améliorent rapidement. Ray Kurzweil prédit qu'un modèle du cerveau de qualité suffisante sera disponible dans un délai similaire à la puissance de calcul requise de 1016 cps vu précédemment. Donc 2025.

Certains projets de recherche étudient la simulation du cerveau à l'aide de modèles neuronaux sophistiqués, mis en œuvre sur des architectures informatiques classiques.

En 2005, le projet "The artificial intelligence system" par Intelligence Realm a mis en œuvre des simulations d'un "cerveau" (avec 1011 neurones) en

temps non réel. Il a fallu 50 jours à un groupe de 27 processeurs pour simuler une seconde d'un modèle similaire au cerveau humain. Pas très efficace, c'est le moins que l'on puisse dire, mais quand même un pas dans la bonne direction.

L'objectif du projet Blue Brain, de l'École Polytechnique Fédérale de Lausanne est de construire des reconstructions et des simulations numériques biologiquement détaillées du cerveau d'un rat, et par la suite du cerveau humain. La recherche consiste à étudier des tranches de tissu cérébral vivant à l'aide de microscopes. Des données sont collectées sur les nombreux types de neurones qui sont utilisés pour construire des modèles biologiquement réalistes de réseaux de neurones. Les objectifs du projet sont d'acquérir une compréhension complète du cerveau et de permettre un meilleur développement sur les traitements des maladies du cerveau.



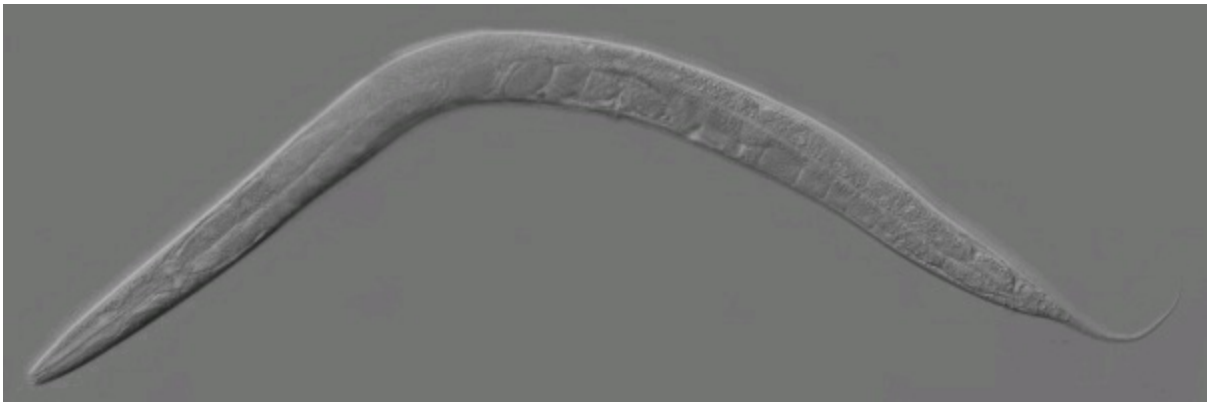

Technique pour "numériser" un cerveau à partir d'un scan IRM

©Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, et al CC BY 3.0

Les chercheurs ont déjà démontré que cette approche peut fonctionner en imitant le cerveau d’un ver de 1mm. Le projet OpenWorm a cartographié les connexions entre les 302 neurones du ver C.elegans et les a simulées dans un logiciel. Le but ultime du projet est de simuler complètement ce ver en tant qu’organisme virtuel. Récemment, la simulation du cerveau du ver a été intégrée dans un simple robot. La simulation n’est pas exacte, mais le comportement du robot est déjà impressionnant étant donné qu'aucune instruction n'a été programmée dans ce robot. Tout ce qu'il a, c'est un réseau de connexions imitant celles du cerveau d'un ver. Alors certes,

302 neurones c’est encore très loin des 100 milliards que contient notre cerveau, mais en prenant en compte une croissance exponentielle, selon les dires de Ray Kurzweil, nous pourrions avoir simulé le cerveau humain aux alentours de 2030.

132



Les 302 neurones du ver c. elegans ont été modéliser sur ordinateur par l'équipe d'OpenWorm ©Kbradnam CC BY-SA 2.5

Une critique fondamentale de la simulation du cerveau concerne le problème du manque d’incarnation. Car la relation au corps est considérée comme un aspect essentiel de l'intelligence humaine. De nombreux chercheurs pensent que cette incarnation est nécessaire pour fonder une véritable intelligence générale. Si cette position est correcte, tout modèle cérébral pleinement fonctionnel devra être incarné dans un corps robotique par exemple. Ben Goertzel propose d’incarner la simulation du cerveau dans un corps virtuel, mais il n’est pas évident que cela soit suffisamment réaliste. Il est d’ailleurs à l’origine de la construction de Sophia, l’un des humanoïdes les plus réalistes aujourd’hui. Si l’on reprend l'expérience du projet OpenWorm, on peut imaginer qu’à l’horizon 2030-2040, une simulation d’un cerveau humain soit incorporée dans un robot humanoïde extrêmement réaliste, ce qui pourrait donner naissance à une sorte d'être hybride humain/machine. Une nouvelle forme de vie et d’intelligence générale.

**Simuler la sélection naturelle**

Si la simulation du cerveau humain est trop compliquée, peut être que l'on peut se pencher sur la reproduction des conditions qui ont menés à l'émergence d'une intelligence générale comme la nôtre. C'est à dire en reproduisant les processus darwiniens à l'oeuvre dans la sélection naturelle et l'évolution des organismes.

Reproduction, sélection, mutation, survie.

Ce champ d'études s'appelle l'informatique évolutive. La manière classique de programmer consiste à écrire du code informatique précis en ayant un objectif spécifique. L'informatique évolutive utilise une approche différente. Cela commence par des centaines de milliers de morceaux de code assemblés au hasard. Chacun de ces codes est testé pour voir s'il atteint l'objectif requis. Et bien sûr, la plupart du code est inutile, car il est généré aléatoirement. Mais certains morceaux de code sont un peu meilleurs que d'autres. Ces morceaux sont ensuite reproduits dans une nouvelle génération de code, qui inclut davantage de copies.



Cependant, la génération suivante ne peut pas être une copie identique de la première. Les nouveaux codes doivent changer soit à travers une mutation ou bien ils peuvent être issus de deux codes qui sont coupés en deux et les moitiés échangées -

comme une recombinaison sexuelle de l'ADN. On a donc des codes "parents" et des codes "enfants". Chaque nouvelle génération est ensuite testée pour vérifier son fonctionnement. Les meilleurs morceaux de code sont reproduits dans une autre génération, et ainsi de suite. De cette façon, le code évolue. Au fil du temps, les choses s'améliorent et après plusieurs générations, si les conditions sont favorables, le code peut devenir meilleures que celui de n'importe qu'elle programmeur humain.

En 2018, Dennis Wilson et ses collègues de l'Université de Toulouse ont montré comment l'informatique évolutive peut égaler les performances des machines utilisant le deep learning dans les jeux d'arcade tels que Pong, Breakout et Space Invaders. Ces travaux suggèrent que les algorithmes

évolutionnaires devraient être utilisés aussi largement que ceux basés sur le machine learning et le deep learning.

C'est une approche flexible pouvant être appliquée à un large éventail de problèmes d'apprentissage et d'optimisation. Cependant, les informaticiens essayant de comprendre et d'utiliser ces approches sont maintenant aux prises avec des problèmes très similaires à ceux rencontrés par les biologistes pour comprendre le fonctionnement de systèmes complexes à différentes échelles. Il faut espérer que les connaissances et l'expérience des deux communautés de recherche pourront être combinées de manière fructueuse pouvant amener à l'émergence d'une IA générale.

L'évolution biologique a mis des milliards d'années pour produire les premières formes d'intelligence générale et j'imagine que les chercheurs aimeraient obtenir des résultats plus rapidement. Mais nous avons beaucoup d'avantages sur l'évolution.

Déjà, les processus évolutionnaires peuvent être exécutés sur des échelles de temps extrêmement plus rapide que celle de l'évolution biologique. Donc nous pouvons voir plus rapidement ce qui fonctionne ou non. Ensuite, l'évolution n'a aucun plan prédéfini, aucun objectif, pas même l'intelligence. Or, nous pourrions spécifiquement orienter ce processus évolutif informatique vers un objectif de générer de l'intelligence. Et enfin, pour privilégier l'intelligence, l'évolution doit innover de différentes manières pour faciliter son émergence, comme par exemple réorganiser la façon dont les cellules produisent de l'énergie. Mais en utilisant directement l'électricité comme source d'énergie aux systèmes évolutifs, cela facilite la progression vers l'objectif visé.

Au final, est-ce que reproduire le cerveau humain est la meilleure façon d'obtenir une intelligence générale de niveau humaine ? Peut être pas. Lorsqu'on regarde l'histoire de l'aviation par exemple, le meilleur moyen de concevoir un engin capable 134

de voler n’a pas été de simuler la façon dont les oiseaux vols en créant des oiseaux mécaniques. Ceux qui ont emprunté cette voie ont connu des échecs cuisants. La solution fut de comprendre les fondamentaux de l'aérodynamisme.

Il en va peut-être de même avec la création d’une intelligence artificielle générale.

Tout comme il y a plusieurs façons de “voler”, il y a surement plusieurs façons d’avoir une intelligence générale. Nous allons peut-être nous retrouver avec des systèmes disposant d’une intelligence similaire à la nôtre, mais fonctionnant sur des principes qui ne sont pas comparables à ceux des neurones et synapses de notre cerveau.

**Comment mesurer l'achèvement d’une intelligence artificielle générale ?**

Admettons qu’une équipe de chercheur prétend avoir mis au point une intelligence artificielle générale. La suite logique c’est de tester leur

affirmation avec des outils d'analyse scientifique. Mais c'est plus facile à dire qu'à faire.

Le concept "d'intelligence générale" fait référence à la capacité d'être efficace dans une multitude de domaines. Ou, comme le dit Ben Goertzel, la capacité d'atteindre des objectifs complexes dans des environnements complexes en utilisant des ressources informatiques limitées. Une autre idée souvent associée à l'intelligence générale est la capacité de transférer l'apprentissage d'un domaine à un autre.

135

Ben Goertzel, ici avec son robot d'Hanson robotics, est l'un des leaders dans la recherche sur l'intelligence artificielle générale. ©Web Summit - DSC_5069 CC BY

2.0

Pour illustrer cette idée, considérons quelque chose qui ne compterait pas comme une intelligence générale. Aujourd'hui, les ordinateurs démontrent des performances surhumaines pour certaines tâches, des performances équivalentes aux humains pour d'autres et des performances sous-humaines pour le reste. Si une équipe de chercheurs était en mesure de combiner dans un seul système, un grand nombre des algorithmes les plus performants appartenant à la catégorie des IA limitées, ils disposeraient d'une

"IA fourre-tout" qui sera terriblement mauvaise dans la plupart des tâches, médiocre dans d'autres, et surhumaine dans une poignée de domaine.

C'est au final, un peu la même chose avec les humains. Nous sommes terriblement mauvais ou médiocres dans la plupart des tâches, et biens meilleurs que la moyenne pour quelques tâches seulement. Car on les a étudiés ou pratiqués beaucoup plus. Une autre similitude est que l'IA fourre-tout montrerait probablement des corrélations mesurables entre de nombreuses capacités cognitives similaires, tout comme les humains où l'on retrouve ce concept avec les tests de QI. Si nous donnions à l'IA fourre-tout beaucoup plus de puissance de calcul, celui-ci pourrait l'utiliser pour améliorer ses performances dans de nombreux domaines similaires.

Par contre, cette IA n'aurait pas (encore) d'intelligence générale, car elle n'aurait pas nécessairement la capacité de résoudre des problèmes arbitraires dans des environnements aléatoires, et ne serait pas nécessairement capable de transférer l'apprentissage d'un domaine à un autre.

Bien que la tâche de mesurer si un système artificiel possède une intelligence générale est compliquée, plusieurs personnes ont proposé des tests.

**Le test de Turing :**

Le test de Turing a été proposé par Alan Turing (1912 - 1954) en 1950, mais a reçu de nombreuses interprétations au fil des décennies.

Une interprétation spécifique est fournie par les conditions pour gagner le prix Loebner. Depuis 1990, Hugh Loebner a offert 100 000 dollars au premier programme qui réussit ce test lors d'un concours annuel. Des prix plus modestes sont décernés chaque année aux IA le plus performantes, mais aucun programme n'a encore remporté le prix de 100 000 $.

Les conditions exactes pour gagner ce grand prix ne seront pas définies tant que le 136

programme ne remportera pas le prix "argent" de 25 000 dollars, ce qui n'a pas encore été fait. Cependant, nous savons que les conditions vont probablement ressembler à ceci : un programme gagnera 100 000 $ s'il peut tromper la moitié des juges en leur faisant croire que c'est un humain en interagissant avec eux dans une conversation sans thème précis pendant 30 minutes et en interprétant des données audiovisuelles.

**Le test du café :**

Ben Goertzel suggère un test probablement plus difficile qu'il appelle le "test au café". Cela paraît simple sur le papier : Entrez dans une maison Américaine moyenne et trouvez comment faire du café, notamment en identifiant la machine à café, en déterminant le fonctionnement des

boutons, en trouvant le café dans le placard, etc.

Selon lui, si un robot pouvait le faire sans avoir été programmé, nous devrions peut-être considérer qu'il possède une intelligence générale.

**Le test du robot étudiant :**

Ben Goertzel, toujours lui, propose une autre mesure encore plus complexe, le

“test du robot étudiant”. Lorsqu'un robot pourra s'inscrire dans une université humaine et suivre des cours de la même manière que les humains et obtenir son diplôme, alors on pourra conclure avec une grande probabilité que ce robot possède une intelligence artificielle générale. Ce qui implique que la durée du test sera de plusieurs mois.

Un ou plusieurs de ces tests peuvent sembler convaincants, mais un regard sur l'histoire permet de nous apprendre une certaine humilité. Il y a des décennies, plusieurs scientifiques de premier plan en intelligence artificielle semblaient penser que la performance aux échecs pouvait représenter un exploit digne d’une IA générale.

En 1976, le mathématicien I.J. Good (1916 - 2009) a affirmé qu’un programme battant un champion humain aux échecs était un bon indicateur d’une IA générale.

Mais les machines ont dépassé les meilleurs joueurs d’échecs humains en 1997, ce qui n’a pas été synonyme d’intelligence générale, mais simplement de force brute informatique et des algorithmes très bien codés. La victoire de DeepMind au jeu de Go est bien plus impressionnante que les échecs, mais cela n’est toujours pas une preuve d’une intelligence générale.

Le succès surprenant des voitures autonomes peut offrir une autre leçon d'humilité. Un scientifique dans les années 1960 aurait peut-être pensé qu'une voiture autonome aussi performante que celle que l’on possède en 2019 serait un signe d’une IA générale. Après tout, une voiture autonome doit agir avec une grande autonomie, à grande vitesse, dans un

environnement extrêmement complexe, dynamique et 137

incertain dans le monde réel. Au lieu de cela, Google a construit sa voiture sans conducteur avec une série de technique qu'il n'était probablement pas imaginable dans les années 1960 - par exemple, en cartographiant avec une grande précision presque toutes les routes, autoroutes et parkings de la planète avant de construire sa voiture sans conducteur. Et le résultat n'est pas une intelligence générale.

Les opinions varient sur la question de savoir si l'intelligence générale artificielle est à prévoir sur le court ou long terme. Herbert A. Simon (1916 - 2001), pionnier de l'intelligence artificielle, écrivait en 1965 : " *Les machines seront capables, dans vingt ans, de faire tout le travail qu'un homme peut faire*". Cependant, cette prédiction ne s'est pas réalisée. Paul Allen (1953 - 2018), cofondateur de Microsoft, estimait qu'une telle intelligence était improbable au 21e siècle, car elle exige des percées fondamentalement imprévisibles et une compréhension scientifique approfondie de la cognition. Le chercheur Alan Winfield a déclaré que le fossé entre l'informatique moderne et l'intelligence artificielle de niveau humain était aussi large que celui existant entre le vol spatial actuel et un vol spatial plus rapide que la vitesse de la lumière. Selon quatre sondages menés en 2012 et 2013, la moyenne parmi les experts sur le moment où l'IA générale arriverait était entre 2040 et 2050. Mais avec des extrêmes allant de 2020 à 2100. Ce sondage n'a pas un poids très élevé dans la mesure où historiquement, les experts d'un domaine ont souvent manqué de flair lorsqu'il s'agit de prédire le développement futur d'une technologie.

138

3. Demain : Intelligence artificielle générale

La question de la conscience

Admettons que nous avons réussi à concevoir une intelligence artificielle générale.

Elle est capable de parler naturellement, comprendre son environnement,

résoudre des problèmes complexes, apprendre dans tous les domaines, raisonner, penser, composer des symphonies. Bref, il n'y a aucun doute sur le fait qu'elle possède une intelligence générale. La question qui se pose ensuite c'est de savoir si quelque de chose de spéciale se passe à l'intérieur de ce système. La lumière est-elle allumée ou éteinte ?

Ou pour le dire plus clairement, possède-t-elle une subjectivité propre que l'on appelle la conscience de soi ?

Les chercheurs sont divisés sur la question de savoir si la conscience peut émerger dans un système artificiel complexe. Il existe également un débat sur le fait de savoir si les machines pourraient ou devraient être qualifiées de "conscientes" au même titre que les humains, et certains animaux. Ou est-ce que la conscience d'une machine est différente ? Certaines de ces questions ont avoir avec la technologie, d'autres avec ce qu'est réellement la conscience.

Car ce mot "Conscience" fait office de controverse dans le milieu de la recherche sur l'intelligence artificielle. Il y a soit des chercheurs qui n'ont pas de temps à perdre avec cette question, d'autres qui y pensent plus ou moins de manières récréatives, et bien sûr, certains philosophes comme Sam Harris, Nick Bostrom ou David Chalmer qui sont préoccupés par les implications éthiques de machine consciente. David Chalmer fait d'ailleurs figure de référence contemporaine sur la question de la conscience. Le philosophe australien est connu pour avoir souligné les deux problèmes sur la recherche de la conscience. Les problèmes faciles et difficiles ("Easy problem of consciousness" et "Hard problem of consciousness").

David Chalmer est un des philosophes contemporain de reference lorsqu'il s'agit des questions sur la conscience. ©Zereshk CC BY 3.0

Il n'y a pas qu'un problème de la conscience. "Conscience" est un terme ambigu, faisant référence à de nombreux phénomènes différents. Chacun de ces phénomènes doit être expliqué, mais certains sont plus faciles à expliquer que d'autres. Au début, il est utile de diviser les problèmes de conscience en "faciles" et "difficiles". Les problèmes faciles de la conscience sont ceux qui semblent directement sensibles aux méthodes classiques des sciences cognitives, selon lesquelles un phénomène est 140

expliqué en termes de traitement de l'information par les différentes parties du cerveau incluant neurones, synapses, cellules gliales. Les problèmes difficiles sont ceux qui semblent résister à ces méthodes.

Pour faire simple, le problème facile de la conscience tente d'expliquer ce qui fait que nous avons des expériences du monde extérieur. Par exemple, dans le cas de la vision, une information visuelle arrive dans notre rétine, puis voyage le long des nerfs optiques et cette information est traitée par le cerveau pour produire une représentation de ce qui est vu par l'oeil.

Le problème difficile de la conscience tente d'expliquer comment et pourquoi on ressent intérieurement les expériences du monde extérieur. Il est indéniable que certains organismes sont des "sujets d'expérience". Mais la question de savoir comment ces systèmes sont des sujets d'expérience est un mystère. Comment se fait-il que lorsque nos systèmes cognitifs s'engagent dans le traitement de l'information visuelle et auditive, nous ayons une expérience subjective visuelle ou auditive. Un organisme est conscient s'il y a quelque chose qui fait d'être cet organisme. Parfois, des termes tels que "conscience phénoménale" et "qualia" sont utilisés ici pour décrire la subjectivité d'une expérience.

Une des réponses possibles c'est que la sélection naturelle a favorisé les expériences conscientes subjectives, car elles permettent une meilleure adaptation à l'environnement et donc augmente les chances de survie d'un organisme. Mais ce n'est pas un argument évident à défendre, car un organisme qui ne possède pas d'expérience subjective peut très bien s'adapter de manière optimale à son environnement. Comme c'est le cas avec les virus ou bactéries. Et la robotique nous montre aussi que le manque de subjectivité n'empêche pas un système d'évoluer dans

un environnement.

Si on prend l'exemple d'une voiture autonome. Elle possède des dizaines de capteurs afin d'analyser l'environnement dans lequel elle se trouve. Lorsqu'une information visuelle est captée par une caméra, elle est traitée puis génère une réaction, par exemple "freiner". À aucun moment, la voiture n'a eu une expérience consciente vis-à-vis de l'information visuelle. Il n'y a rien qui fait d'être cette voiture.

Il n'empêche que son comportement peut-être cohérent et elle est capable d'accomplir des objectifs complexes comme conduire d'un point A à un point B en toute sécurité.

Ce qui la rend "intelligente" selon la définition que l'on donne dans la première partie.

Pourquoi les humains et un grand nombre d'animaux possèdent une subjectivité des expériences extérieures, ce qui crée un monde intérieur ? Nous pourrions très bien 141

être des "zombies philosophiques". Terme utilisé pour définir un être hypothétique qui, de l'extérieur, est impossible à distinguer d'un être humain normal, mais qui manque d'expérience subjective consciente. Si tous les organismes n'étaient ni plus ni moins que des zombies philosophies, ou de simples machines, ils auraient tout autant de capacité à évoluer dans leur environnement et survivre.

Appartenant pendant longtemps à la philosophie et aux religions puis aux neurosciences, la question de la conscience est désormais associée à la recherche sur l'intelligence artificielle. Si nous ne savons pas comment la conscience émerge ni comment la détecter, alors comment pourrons-nous être sûrs que les intelligences artificielles de demain possèdent ou non des expériences subjectives. Surtout si nous arrivons à concevoir une intelligence artificielle générale.

Autant des progrès significatifs sont faits en intelligence artificielle, aucun n'a été fait dans le développement de la conscience artificielle. Certains chercheurs pensent que la conscience est une caractéristique qui émergera à

mesure que la technologie se développe. Certains pensent que la conscience implique l'intégration de nouvelles informations, de stocker et de récupérer des anciennes informations et de les traiter de manière intelligente. Si cela est vrai, les machines deviendront forcément conscientes.

Car elles pourront collecter bien plus d'informations qu'un humain, les stocker virtuellement sans limites, accéder à de vastes bases de données en quelques millisecondes et les calculer de façon bien plus complexe que notre cerveau. Le résultat sera logiquement l'émergence d'une conscience subjective artificielle. Et peut être même encore plus fine et profonde que la nôtre.

On peut considérer que plus un organisme est complexe, plus sa subjectivité consciente est grande. Ce qui explique pourquoi nous avons un éventail d'expériences conscientes subjectives plus large et profond qu'un chien, une hirondelle ou un serpent. Par conséquent, mettre au point une intelligence artificielle plus complexe, entraînera peut être l'émergence d'une entité possédant une conscience largement supérieure à la nôtre.

Un autre point de vue sur la conscience provient de la physique quantique, qui explique les lois de la physique à l'échelle de l'infiniment petit. Selon l'interprétation de Copenhague, la conscience et le monde physique sont des aspects complémentaires de la même réalité. Lorsqu'une personne observe ou expérimente un aspect du monde physique, son interaction consciente provoque un changement. L'interprétation de Copenhague considère la conscience comme une chose qui existe par elle-même dans l'univers - même si cela nécessite un cerveau pour qu'elle existe dans le monde physique. Ce point de vue était populaire parmi les pionniers de la physique quantique tels que Niels Bohr, Werner Heisenberg et Erwin Schrödinger.



La vision opposée est que la conscience émerge de la biologie, tout comme la biologie elle-même émerge de la chimie qui, à son tour, émerge de la physique. Il s'agit de l'opinion de beaucoup de neuroscientifiques selon laquelle les processus de l'esprit sont identiques

aux états et aux processus du cerveau. Ces conceptions modernes de la conscience vue par la physique quantique ont des parallèles avec les anciennes philosophies antiques. L'interprétation de Copenhague ressemble aux hypothèses présentes dans la philosophie indienne, dans laquelle la conscience est la base fondamentale de la réalité.

La vision émergente de la conscience, en revanche, est assez similaire au bouddhisme. Bien que le Bouddha ait choisi de ne pas aborder la question de la nature de la conscience, ses disciples ont déclaré que l'esprit et la conscience découlent du vide ou du néant.

Mais si on se place au niveau de la physique, un être humain n'est qu'un ensemble d'atome et de quarks arrangés d'une certaine manière. Qu'est-ce qui fait qu'un certain arrangement d'atome et de quarks est conscient, et un autre arrangement ne l'est pas ?

Aujourd'hui, si on se pose la question : "Est-ce que cela fait quelque chose d'être AlphaGo, une voiture autonome ou Google Home ? La réponse est probablement : non. Car il n'y a aucune subjectivité. La lumière est éteinte. Mais demain ? Avec des systèmes artificiels bien plus complexes, ayant la puissance du calcul du cerveau humain et capable d'accomplir autant de tâches qu'un humain, ce ne sera peut être pas aussi évident de conclure qu'ils n'ont aucune conscience. Et cela pourrait entraîner de nombreux problèmes éthiques.



## 3. Demain : Intelligence artificielle générale

### Les problèmes éthiques à venir

L'arrivée d'une intelligence artificielle générale posera de nombreuses questions éthiques et les chercheurs ont bien compris les implications pour le futur de l'humanité. Le "Future of Life Institute" (Institut pour le Futur de la Vie) fondé par le cosmologiste Max Tegmark, a organisé une conférence en janvier 2017 intitulé : The Asilomar Conference on Beneficial AI (la conférence d'Asilomar sur l'IA bénéfique).

Plus de 100 penseurs et chercheurs en intelligence artificielle, économie, en droit, en éthique et en philosophie se sont rencontrés lors de la conférence pour aborder et formuler un ensemble de lignes directrices pour la création d'intelligences artificielles bénéfiques. Il s'agit des 23 principes d'Asilomar sur l'IA :

**Les 23 principes d'Asilomar sur l'intelligence artificielle**

1) Objectif de ces recherches : Le développement de l'IA ne doit pas servir à créer une intelligence sans contrôle mais une intelligence bénéfique.

2) Investissements : Les investissements dans l'IA doivent être soutenus par le financement de recherches visant à s'assurer de son usage bénéfique, qui prend en compte des questions épineuses en matière d'informatique, d'économie, de loi, d'éthique et de sciences sociales.
Quelques exemples : Comment rendre les futures IA suffisamment solides pour qu'elles fassent ce qu'on leur demande sans dysfonctionnement ou risque d'être piratées ? ou encore comment améliorer notre prospérité grâce à cette automatisation tout en maintenant les effectifs humains ?

3) Relations entre les scientifiques et les législateurs : Un échange constructif entre les développeurs d'IA et les législateurs est souhaitable.

4) Esprit de la recherche : Un esprit de coopération, de confiance et de transparence devrait être entretenu entre les chercheurs et les scientifiques en charge de l'IA.

5) Éviter une course : Les équipes qui travaillent sur les IA sont encouragées à coopérer pour éviter des raccourcis en matière de standards de sécurité.

6) Sécurité : Les IA devraient être sécurisées tout au long de leur existence, une caractéristique vérifiable et applicable.

7) Transparence en cas de problème : Dans le cas d'une blessure provoquée par une IA, il est nécessaire d'en trouver la cause.

8) Transparence judiciaire : Toute implication d'un système autonome

dans une décision judiciaire devrait être accompagnée d'une explication satisfaisante contrôlable par un humain.

144

9) Responsabilité : Les concepteurs et les constructeurs d'IA avancées sont les premiers concernés par les conséquences morales de leurs utilisations, détournements et agissements. Ils doivent donc assumer la charge de les influencer.

10) Concordance de valeurs : Les IA autonomes devraient être conçues de façon à ce que leurs objectifs et leur comportement s'avèrent conformes aux valeurs humaines.

11) Valeurs humaines : Les IA doivent être conçues et fonctionner en accord avec les idéaux de la dignité, des droits et des libertés de l'homme, ainsi que de la diversité culturelle.

12) Données personnelles : Chacun devrait avoir le droit d'accéder et de gérer les données le concernant au vu de la capacité des IA à analyser et utiliser ces données.

13) Liberté et vie privée : L'utilisation d'IA en matière de données personnelles ne doit pas rogner sur les libertés réelles ou perçue des citoyens.

14) Bénéfice collectif : Les IA devraient bénéficier au plus de gens possible et les valoriser.

15) Prospérité partagée : La prospérité économique permise par les IA devrait être partagée au plus grand nombre, pour le bien de l'humanité.

16) Contrôle humain : Les humains devraient pouvoir choisir comment et s'ils veulent reléguer des décisions de leur choix aux IA.

17) Anti-renversement : Le pouvoir obtenu en contrôlant des IA très avancées devrait être soumis au respect et à l'amélioration des processus civiques dont dépend le bien-être de la société plutôt qu'à leur détournement.

18) Course aux IA d'armement : Une course aux IA d'armement

mortelles est à éviter.

19) Avertissement sur les capacités : En l'absence de consensus sur le sujet, il est recommandé d'éviter les hypothèses au sujet des capacités maximum des futures IA.

20) Importance : Les IA avancées pourraient entraîner un changement drastique dans l'histoire de la vie sur Terre, et doit donc être gérée avec un soin et des moyens considérables.

21) Risques : Les risques causés par les IA, particulièrement les catastrophiques ou existentiels, sont sujets à des efforts de préparation et d'atténuation adaptés à leur impact supposé.

22) Auto-développement infini : Les IA conçues pour s'auto-développer à l'infini ou s'auto-reproduire, au risque de devenir très nombreuses ou très avancées rapidement, doivent faire l'objet d'un contrôle de sécurité rigoureux.

23) Bien commun : Les intelligences surdéveloppées devraient seulement être développées pour contribuer à des idéaux éthiques partagés par le plus grand nombre et pour le bien de l'humanité plutôt que pour un État ou une entreprise.

Ces 23 principes sont largement acceptés par la communauté scientifique ce qui représente une bonne nouvelle pour la sécurité du développement de l'intelligence 145

artificielle.

Le physicien Max Tegmark est un des porteurs des 23 principes d'Asilomar grâce au Future of Life Institute. ©Bengt Oberger CC-BY-SA-4.0

**L’éthique des machines**

L'éthique des machines (ou la morale des machines) est le domaine de recherche qui concerne la conception d'agents moraux artificiels, que ce soit des robots ou des programmes qui se comportent moralement ou comme s'ils étaient doués de valeurs morales.

En 2009, des universitaires et des experts ont assisté à une conférence organisée par l'Association pour l'avancement de l'intelligence artificielle (Association for the Advancement of Artificial Intelligence) afin de discuter de l'impact potentiel des robots et des ordinateurs dans l'hypothèse selon laquelle ils pourraient devenir autonomes et capables de prendre leurs propres décisions. Ils ont noté que certaines machines avaient acquis diverses formes de semi-autonomie, en étant capables notamment de trouver elles-mêmes des sources d'énergie et de pouvoir choisir indépendamment les cibles à attaquer.

Cependant, il existe une technologie en particulier qui pourrait véritablement concrétiser la possibilité de robots dotés de compétences morales. Dans un article, Nayef Al-Rodhan mentionne le cas des puces neuromorphes, qui visent à traiter les informations de la même manière que les humains, c'est à dire de manière non linéaire et avec des millions de neurones artificiels interconnectés. Les robots intégrés à la technologie neuromorphique pourraient apprendre et développer des 146

connaissances de manière humaine. Cela pose la question de l'environnement dans lequel de tels robots apprendraient du monde et de la moralité dont ils hériteraient, ou s'ils développeraient également des "faiblesses" humaines : égoïsme, attitude prosurvie, voire même s'ils pourraient avoir des opinions racistes, discriminatoire ou des biais de confirmation (qui consiste à privilégier les informations confirmant ses idées préconçues ou ses hypothèses sans considération pour la véracité de ces informations et/ou à accorder moins de poids aux hypothèses et informations jouant en défaveur de ses conceptions.)

Alors que les algorithmes du machine learning sont utilisés dans des domaines aussi variés que la publicité en ligne, le prêt bancaire et la condamnation pénale, avec la promesse de fournir des résultats plus objectifs et fondés sur les données, il a été prouvé qu'ils peuvent être source d'inégalités sociales et de discrimination. Par exemple, une étude réalisée en 2015 a révélé que les annonces Google d'emploi à revenu élevé étaient moins affichées aux femmes et une autre étude a révélé que le service de livraison en 1 jour ouvré d'Amazon n'était pas toujours disponible dans les quartiers noirs, pour des raisons qu'Amazon ne pouvait pas expliquer, car ils ne savent pas tout ce qu'il se passe sous le capot de leurs algorithmes.

Afin d'éviter d'augmenter les taux d'emprisonnement déjà élevés aux États-Unis, des tribunaux américains ont commencé à utiliser un logiciel d'évaluation des risques lorsqu'ils décident des libérations sous caution. Ces outils analysent l'historique des accusés et d'autres paramètres. Un rapport de 2016 de ProPublica a analysé les risques de récidive calculés à l'aide de l'un des outils les plus couramment utilisés, le système Northpointe COMPAS, et a examiné les résultats sur deux ans. Il en ressort que 61%

seulement des personnes considérées à risque élevé ont effectivement commis d'autres infractions et que les accusés afro-américains étaient beaucoup plus susceptibles d'être jugé à haut risque que les accusés blancs.

Considérés comme un biais algorithmique, les systèmes informatiques agissent sur les préjugés des personnes qui entrent et structurent leurs données.

En mars 2018, dans le but de répondre aux préoccupations croissantes concernant l'impact des machines sur les droits de l'homme, le Forum sur l'économie mondiale et le Conseil pour l'avenir des droits de l'homme dans le monde ont publié un article contenant des recommandations détaillées sur les meilleurs moyens de prévenir les biais algorithmiques.

Les recommandations du Forum économique mondial sont les suivantes :

- Intégration active: le développement et la conception d'applications de machine 147

learning doivent rechercher activement une diversité dans les résultats, en particulier dans les normes et les valeurs d'une population spécifique affectée par le résultat des algorithmes.

- Équité: les personnes impliquées dans la conceptualisation, le développement et la mise en œuvre de systèmes de machine learning doivent déterminer quelle définition de l'équité s'applique le mieux à leur contexte et à leur application.

- Droit à la compréhension: L'implication du machine learning dans la prise de décision qui affecte les droits individuels doit être divulguée. Les algorithmes doivent être en mesure de fournir une explication de leur prise de décision qui soit compréhensible pour les utilisateurs et pouvant être révisée par une autorité humaine compétente. Lorsque cela est impossible et que des droits sont en jeu, les responsables de la conception, du déploiement et de la réglementation des algorithmes doivent se demander s'ils doivent être utilisés ou non.

- Accès à la correction: les responsables, les concepteurs et les développeurs

de systèmes de machine learning sont chargés d'identifier les conséquences potentiellement négatives de leurs systèmes sur les droits de l'homme. Ils doivent offrir des voies de recours aux personnes touchées négativement par la décision d'algorithmes et mettre en place des processus permettant de remédier aux discriminations.

**La roboéthique**

Ce sous-domaine de l'éthique des technologies concerne les problèmes éthiques concernant les robots. Notamment la façon dont ils peuvent représenter des menaces pour la liberté, le droit et la survie humaine. Mais il y a également un autre angle d'approche vis-à-vis de la roboéthique : comment traitons-nous les robots ? Et par extension l'intelligence artificielle ?

La question paraît ridicule aujourd'hui, mais nous avons exploré précédemment qu'un système artificiel pourrait acquérir une intelligence générale. Dès lors, l'idée selon lequel les humains devraient avoir des obligations morales envers leurs machines, similaires aux droits de l'homme ou aux droits des animaux, sera à considérer. Il a été suggéré que les droits des robots, tels que le droit d'exister et d'accomplir sa propre mission, soient liés au devoir de servir l'homme, par analogie avec le lien entre les droits de l'homme et les devoirs de l'homme devant la société, notamment le droit à la vie et à la liberté, la liberté de pensée et d'expression et l'égalité devant la loi.

La question de la souffrance et de la douleur est un sujet complexe qui a des 148

implications radicales. La série Westworld est clairement une référence pertinente sur cette question. Pour résumer rapidement, Westworld raconte l'histoire d'un nouveau genre de parc d'attractions se situant dans un futur proche. Pour un prix quotidien très élevé, vous pouvez vivre des aventures diverses dans une recréation du Far West en côtoyant des androïdes. Une sorte de jeu de rôle géant sans conséquence puisque les humains ne peuvent pas être gravement blessés. Ce parc ouvre donc la porte à la débauche et fait ressortir chez certains leur part d'ombre lorsqu'ils violent et torturent les androïdes sans aucune retenue morale.

Du coup, la série soulève plusieurs questions morales concernant nos actes vis-à-

vis de créatures si perfectionnées qu'elles sont semblables aux humains. En d'autres termes, est-ce que je dois me sentir mal si je torture un robot ? Où dois-je me retenir de toute barbarie ?

Ça peut paraître assez trivial, mais c'est une question compliquée lorsqu'on commence à déplacer la frontière de ce que l'on considère comme "moral". Tout d'abord, imaginer que vous vous trouviez dans la chambre d'un saloon, au 19e siècle.

Une charmante jeune courtisane (ou courtisant en fonction de vos goûts) arrive et vous passez de très longues minutes à utiliser son corps pour votre plaisir. Vous laissez complètement aller vos fantasmes les plus inavouables et pour une raison inconnue, vous décidez de lui couper la tête ! Soudain, vous vous réveillez en sursaut.

Votre conjoint dort à côté de vous. Ce n'était qu'un rêve ! Est-ce que vous avez transgressé vos valeurs morales ? Votre intuition vous dit probablement que non.

Après tout, vous étiez inconscient, et vous n'avez pas réellement trompé votre conjoint. Encore moins blessé votre partenaire imaginaire. Au pire, vous vous sentez un peu bizarre, mais ce n'était qu'un rêve ! Donc en quoi est-ce différent de

"Westworld" ?

Passons maintenant à la série de jeux vidéo "GTA – Grand Theft Auto". Le jeu nous donne une liberté d'action très vaste où il est possible de voler des voitures, écraser des piétons, personne ne se retient de tuer sans remords les pauvres personnages numériques ! Cela va sans dire ! Devrions-nous ressentir plus de compassion ? Est ce que les dizaines de millions de joueurs sont des êtres immoraux ?

Que se passera-t-il lorsque dans GTA 10, les personnages numériques seront si

“humains”, que cela en sera troublant ? En quoi est-ce différent de “Westworld” ?

Dans la même lignée, est-ce que c’est immoral d’insulter Google Home, Siri ou Alexa ? Comme pour le cas du rêve, la réponse intuitive c’est que ce n’est pas immoral. Mais est-ce que ce sera la même réponse dans quelques années ? Autrement dit, est-ce que c’est fondamentalement mal de torturer des automates ou personnages de jeux vidéos ayant suffisamment de complexité pour passer le test de Turing et 149

ayant une intelligence générale ?

Par contre, une chose qui n’a pas besoin d’être débattue moralement, c’est de mettre un chat dans un micro-onde et le regarder avec cruauté. C’est immoral, on est tous 100% d’accord là-dessus j’espère ! Mais en quoi est-ce différent de “Westworld”

?

La question, au final, est de se demander où se situe la frontière entre un acte immoral sur un animal, et un acte immoral sur un personnage artificiel (virtuel ou physique). Y a-t-il vraiment une frontière au final ?

Chez les humains, dans une situation donnée, la douleur et le plaisir informent bien plus nos processus décisionnels que la connaissance rationnelle de cette situation.

Même si bien sûr, la pensée rationnelle peut aider à clarifier le contexte. Du coup, même si un robot peut traiter une multitude d’information complexe dans son environnement, il n’aura aucun intérêt de stresser à la vue d’une machette sur le point de lui trancher un bras, car il n'a aucune peur d’avoir mal. En revanche, on peut le programmer pour prétendre ressentir du stress voir de la souffrance, et donc avoir un comportement crédible (c’est d’ailleurs le cas dans la série). Mais il y a une différence entre faire semblant d’avoir mal, et véritablement ressentir la douleur. D’ailleurs, les acteurs sont les mieux placés pour en parler, eux qui se mettent dans des situations parfois horribles dans des films.

C’est comme expliquer à un aveugle ce qu’est la couleur rouge. Vous pourrez lui détailler toutes les propriétés physiques de la couleur rouge et lui faire un exposé sur le spectre de la lumière visible, c’est seulement en retrouvant la vue qu’il ressentira ce qu’est véritablement la couleur rouge.

Pour reprendre le cas du rêve mentionné précédemment. On ne peut pas être tenu moralement responsable de ce que l’on fait à un être sorti tout droit de notre imagination. Aussi réaliste qu’il puisse être. Tout comme pour un personnage virtuel, et par extension, un robot. Ce n’est pas parce qu’un androïde est similaire à un humain que nous devons lui accorder des droits, privilèges et le considérer comme un humain. Donc ce qui fait la différence entre l’immoralité d’un chat dans un micro-onde et la cruauté dans Westworld, n’a rien à voir avec la conscience de soi, l’anthropomorphisme, la mémoire, la sophistication de la créature ou tous autres émotions complexes. C’est une question de ressentir la souffrance. Attention à ne pas confondre souffrance et douleur. La douleur n’est qu’une réaction nerveuse, un réflexe biologique qui se passe dans le cerveau après un stimulus sensoriel et qui empêche une créature de se causer des dommages irréversibles. C’est un outil utile que la nature a trouvé pour accentuer l’instinct de conservation. C’est ce qui fait que lorsque vous posez votre main sur un radiateur bouillant, vous l’enlevez aussitôt par réflexe.

150

La souffrance est un concept bien plus profond qui implique une expérience cognitive et émotionnelle résultant d’une douleur physique ou psychologique subjective. On sait que les animaux peuvent souffrir, donc leur faire du mal est fondamentalement immoral et condamnable par la même occasion. En revanche, sur le papier, un système d’intelligence artificielle n’est ni plus ni moins qu’un gadget perfectionné du 21e siècle. Et même si, dans quelques décennies, ces gadgets deviennent indissociables à un humain, ça ne change pas le fait qu’ils resteront des outils, des objets et des biens matériels. Donc si vous coupez le bras d’un de ces androïdes du futur, ce n’est fondamentalement pas immoral. Ce sera simplement un beau gâchis de détruire une technologie sophistiquée et probablement très cher.

Il faut reconnaître que nous avons tendance à attacher beaucoup d'importance à l'anthropomorphisme. Ce qui nous ressemble génère de l'empathie. C'est un processus évolutionnaire profondément ancré en nous. Pour cette raison, des androïdes réalistes pourraient un jour se voir accorder certains droits et protections législatives. En revanche, il n'est pas du tout rationnel de se lamenter des horreurs que ces robots pourraient subir. Que ce soit des robots sexuels ou des souffre-douleurs dans un parc d'attractions. Ni de craindre que leurs souffrances les amènent à réclamer justice en nous attaquant. En l'absence de réelle souffrance, il n'est pas sûr d'affirmer qu'un ensemble de composants mécaniques sophistiqué engendrera des réactions émotionnelles comme le ressentiment, la peur, la colère et la vengeance.

Chez les humains, ces réactions émotionnelles sont dues à bien plus que de simples traitements d'information et puissance de calcul. Il y a des millions d'années d'évolutions derrières.

Si on en croit les déclarations du Professeur Ford, créateur du parc et des robots dans la série, alors la souffrance est la dernière clé, après la mémoire et l'improvisation, pour atteindre la conscience. C'est une théorie de Dostoïevski, philosophe russe : la souffrance est l'unique origine de la conscience. Car il n'est pas possible d'être conscient en étant enfermé dans une bulle où rien de mauvais n'arrive et ne jamais se poser de question sur son existence. La souffrance pousse l'individu à se poser la question "Pourquoi est ce que ça m'arrive ". C'est d'ailleurs un des arguments majeurs mis en avant par les organismes de protection des animaux. Ils souffrent, donc ils sont conscients, donc arrêter de leur faire du mal !

Il faudra donc se demander si les futurs androïdes peuvent ressentir la souffrance, afin de délimiter clairement la frontière morale. Et ce sera compliqué, car comment peut-on réellement savoir si une machine expérience une souffrance comme l'humiliation, la crainte, voir même le deuil, ou bien si elle ne fait que délivrer une réponse qu'elle a apprise en analysant des donnés ? D'un point de vue technologique, il est possible de connecter une partie du corps d'un robot, et de programmer : "Si la

pression est trop forte, envoyer une réponse négative". En simplifiant, cela revient à programmer la douleur. Mais de là à dire que cette douleur évoluera en souffrance, comme chez les animaux, c'est un raccourci loin d'être évident. Peut être qu'ils ressentiront la douleur et diront : "J'ai mal, je souffre, stoppe", mais cela n'aura aucun sens profond pour eux. Ce ne sera qu'une réponse automatique.

Une situation similaire à "Westworld" pourrait arriver très rapidement dans nos sociétés avec des endroits où l'on peut faire ce que l'on veut sur des androïdes réalistes. Même si ce sera considéré comme des parcs pour psychopathe. Il y aura alors un débat semblable à celui qui touche régulièrement les jeux vidéos violents.

Peut être que ces derniers participent à une brutalisation générale de la société, mais il se peut également qu'ils empêchent la violence dans le monde réel en fournissant une alternative virtuelle. Le débat ira bien plus loin que ça en affirmant que c'est immoral de torturer, violer ou tuer un androïde. Des associations pour la protection des robots verront peut-être même le jour. La problématique ne concernera pas que des androïdes d'ailleurs, mais également des simulations virtuelles et les assistants virtuels.

Donc si tout système éprouvant de la souffrance est conscient, alors cette définition laisse ouverte la possibilité que certaines IA puissent aussi être conscientes, même si elles existent uniquement en tant que programme et qu'elles ne sont pas connectées à des capteurs ou des corps robotiques. Du coup, il est difficile de ne pas se soucier de l'immense souffrance que l'on pourrait affliger sans même le savoir à des entités intelligentes. Comme Yuval Noah Harari le souligne dans son livre Homo Deus : "Si un scientifique veut mettre en avant l'argument que les expériences subjectives ne sont pas pertinentes, son défi consiste à expliquer pourquoi la torture et le viol sont des actes immoraux en ne faisant référence à aucune expérience subjective". Sans une telle référence, effectivement, il ne s'agit juste que d'un tas de particules élémentaires qui se déplacent selon les lois de la physique. Rien d'immoral.

Mais on sait que les victimes de torture et viol ont des expériences subjectives horriblement négatives et c'est pour cette raison que l'on condamne ces pratiques.

Si une intelligence artificielle peut éprouver de la souffrance, nos obligations morales devront être clairement définies ©Ritchie333 CC-BY-2.0

Le terme “Mind crime” (crime sur esprit) a été popularisé par les chercheurs en intelligence artificielle pour parler des souffrances innommables que nous pourrions infliger à des systèmes artificiels conscients. Que nous le sachions, ou pas. Car si nous générons en 2050, un trillion d’émulations artificielles d’humain dans des simulations virtuelles, il faudra se demander si éteindre son ordinateur résulte en un génocide à l’ampleur astronomique.

Selon moi, le progrès moral de l'espèce humaine est de contribuer à la réduction de la souffrance pour toutes les entités capable d'expérimenter cette expérience subjective négative. Nous avons clairement fait des progrès majeurs dans la réduction de la souffrance au sein de notre espèce. Réduction de la mortalité infantile, des maladies infectieuses, des famines, de la violence, de l’esclavage, du travail forcé, des guerres, des

sans-abris, augmentation de la santé, l'éducation, l'accès à la connaissance, la richesse, le confort, l'espérance de vie et bien d'autres. Par contre, il est difficile d'être aussi positif concernant la réduction de la souffrance chez les autres espèces. On s'est longtemps demandé si les animaux expérimentent la conscience, et donc la souffrance. Mais aujourd'hui, on sait que c'est le cas pour une grande majorité du règne animal. En tous cas, au-dessus d'une certaine échelle. Si bien que face à l'industrialisation de l'élevage intensif notamment, on ne peut que conclure que l'on a engendré une quantité ahurissante de souffrance. Ce qui ternit notre progrès moral.

153

Il nous faudra donc être extrêmement prudents face à la possibilité que les intelligences artificielles générales de demain puissent être conscientes et souffrir. Ou alors nous risquons de créer des véritables enfers et de les peupler par des millions d'esprits non humains. Quelle sorte d'espèce serons-nous alors ?

154

4.

Après demain : Super intelligence artificielle

155

4. Après demain : Super Intelligence Artificielle

L'explosion d'intelligence

Lorsque nous aurons réussi à concevoir une intelligence artificielle qui possède des facultés d'apprentissage similaire à celle d'un humain (Meta apprentissage). C'est à dire capable d'apprendre sans une quantité astronomique de données, mais également d'appliquer cette technique d'apprentissage à plusieurs domaines sans repasser par la case départ, nous aurons alors passé d'IA limitée, à IA générale.

Si on prend l'exemple d'Alpha Zero de Google, et qu'on lui donne la puissance de calcul du cerveau humain et du meta apprentissage comparable aux humains. Que se passe-t-il ?

Il est déjà le meilleur joueur de Go au monde, ainsi qu'aux échecs. Maintenant, admettons qu'on lui demande d'apprendre à jouer au jeu vidéo StarCraft. Résultat : il devient le meilleur du monde en quelques heures. Ensuite, on lui demande d'apprendre à conduire une voiture, à cuisiner, à écrire des livres, à faire des mathématiques, à entretenir une conversation, à composer des symphonies. Bref plus ou moins tout ce qu'un humain peut faire de manière générale. Il va donc devenir aussi performant qu'un humain.

En faite, non … il sera bien plus performant qu'un humain. Pourquoi ?

Si l'on compare les trois grandes révolutions que l'on a vues au premier chapitre (Révolution cognitive, révolution agricole et révolution scientifique), la révolution de l'intelligence artificielle est bien plus significative que la révolution agricole et la révolution scientifique. Elle se rapproche plus de la révolution cognitive par l'ampleur du changement qu'elle permet. Mais on pourrait même supposer que cette quatrième révolution, que nous vivons actuellement, aura encore plus d'ampleur pour homo sapiens.

Effectivement, la transition vers l'ère des machines intelligentes pourrait être bien plus fondamentale que le passage de notre ancêtre commun vers les premiers membres du genre Homo il y a plus de 2,5 millions d'années.

Car pour arriver aux premiers humains, il a fallu faire croître la taille du cerveau de nos ancêtres commun par un facteur d'environ trois. Et quelques modifications dans la structure neurologique permettant un peu plus de capacité cognitive. Donc 156

pour le dire autrement, ce qui différencie fondamentalement Homo Sapiens des gorilles, orangs-outans et chimpanzés, c'est une boite crânienne un peu plus grande et quelques connexions synaptiques en plus, notamment dans le néocortex. C'est cette petite différence qui donne la possibilité à une espèce de marcher sur la Lune, comprendre les lois de la physique, modifier génétiquement d'autres organismes et composer des symphonies, tandis que l'autre utilise un bâton pour cueillir des bananes. Et si cette différence est si petite, c'est au final, car il n'y avait pas vraiment d'autres possibilités. Une différence plus grande, par exemple si la sélection naturelle

"avait voulu" tester un cerveau cinq fois plus grand au lieu de trois fois, les femelles n'auraient tout simplement pas pu mettre au monde leurs progénitures sans que cela se termine en bain de sang (c'est déjà assez difficile comme ça !). Donc pas la meilleure solution pour faire perdurer une espèce. Notre boîte crânienne à la taille maximale la plus proche des limites biologiques. Ce qui fait que notre cerveau ne peut pas devenir plus gros par la simple voie de la sélection naturelle.

Par contre, si on regarde la différence entre un cerveau humain, et ce que le cerveau d'une intelligence artificielle pourrait devenir, on fait face à un gouffre monumental qui ne possède pas les mêmes limites biologiques. Une intelligence artificielle n'a pas à rester confinée dans une boîte crânienne, elle peut être aussi large qu'une planète. Donc ça fait déjà une grande différence dans la puissance de l'intelligence produite.

Notre cerveau ne pouvait pas être plus que 3 fois plus gros que celui de nos ancêtres communs en raison de la taille de la boite crânienne. Mais une IA n'a aucune limite tangible. ©Gaëtan Selle

Mais ce n’est pas tout. Car il existe également d’autres limites biologiques dans le cerveau. La vitesse de communication entre deux neurones n'excède pas 120m/s et l’activité d’un neurone est d’environ 100/200 Hz, c’est à dire jusqu'à 200 fois par seconde. Aujourd’hui, nos transistors opèrent déjà à 3 GHz (trois milliards de fois par seconde) et la vitesse de communication d’un signal informatique est limité par la vitesse de la lumière, 299 792 458 m/s.

157

En termes de mémoire, la biologie est également limitée tandis qu’un ordinateur possède une mémoire de travail beaucoup plus grande (RAM) et une mémoire à long terme (stockage sur disque dur) qui offre à la fois une capacité et une précision bien supérieures à la nôtre. Une intelligence artificielle générale n’oubliera jamais rien dans les millions de téraoctets qu’elle aura emmagasinés. Les transistors informatiques sont également plus précis que les neurones biologiques et sont moins susceptibles de se détériorer (et peuvent être réparés ou remplacés). Les cerveaux humains se fatiguent aussi facilement, alors que les ordinateurs peuvent fonctionner sans interruption, à des performances optimales, 24 heures sur 24, 7 jours sur 7 avec comme seul besoin, de l'électricité.

Mais on pourrait se dire que la force de l’humanité c’est notre intelligence collective. À partir du développement du langage et de la formation de communauté, en passant par l’invention de l’écriture et de l’imprimerie, et qui s’intensifie désormais grâce à des outils comme Internet. L’intelligence collective de l’humanité est l’une des principales raisons pour lesquelles nous avons pu aller aussi loin comparer aux autres espèces sur la planète. Donc nous aurons quand même cet avantage face à une intelligence artificielle générale.

Mais c’est oublié que les ordinateurs sont déjà bien meilleurs que nous à collaborer en réseau. Ils peuvent s'échanger des informations à la vitesse de la lumière. Une intelligence artificielle générale pourrait créer une multitude de copies d’elle même et les envoyer aux quatre coins du net formant un réseau mondial d'IA de sorte que tout ce que les versions apprendront sera instantanément téléchargé à toutes les autres. Bien plus efficace que l’intelligence collective humaine. Le réseau d’IA pourrait

également avoir un seul objectif en tant qu'unité, car il n'y aurait pas nécessairement d'opinions, de motivations et d'intérêts personnels dissidents, contrairement à ce que l'on observe au sein de la population humaine.

Le potentiel d'une intelligence produite sur un substrat informatique dépasse donc largement celui d'une intelligence produite sur un substrat biologique. À tous les niveaux !

C'est pour ces raisons qu'une intelligence artificielle générale, comme Alpha Zero pour reprendre notre exemple plus haut, sera une forme d'intelligence dépassant celle d'un être humain. Une première dans l'histoire de notre espèce. Une intelligence artificielle de niveau humain ne signifie pas qu'elle possède la même intelligence qu'un humain, mais le simple fait de posséder une capacité d'apprentissage générale la rend directement super humaine, en raison de ces avantages par rapport à une intelligence biologique.



Cela pose plusieurs questions concernant notre relation avec une telle intelligence.

Comment l'humanité va-t-elle cohabiter avec une entité plus intelligente ? Comment va-t-on la contrôler ? Sera-t-elle accessible par l'ensemble de l'humanité ou que par une élite, un gouvernement ? Nous nous pencherons sur ces questions dans les chapitres suivants.

Mais l'élément à ne pas oublier, c'est que cette intelligence capable d'apprendre comme un humain sera aussi capable d'apprendre la programmation informatique.

Autrement dit à se modifier elle même. Tout comme elle est devenue meilleure à toutes les tâches qu'elle a apprises, elle deviendra aussi la meilleure programmeuse du monde. Elle pourra se mettre à jour, augmenter ses performances, trouver de nouvelle façon d'accélérer sa vitesse et puissance de calcul. Cette machine super intelligente conçoit alors une machine encore plus performante ou réécrit son propre logiciel pour devenir

encore plus intelligente. Cette machine (encore plus performante) continue ensuite à concevoir une machine encore plus performante, et ainsi de suite jusqu'à atteindre les limites permises par les lois de la physique. On entre alors dans une boucle d'auto amélioration récursive qui engendre une explosion d'intelligence. Terme inventé par le mathématicien Irvin J. Good (1916 - 2009) en 1965. Un autre terme utilisé pour décrire ce même phénomène c'est la singularité technologique.

Le mot singularité est issu de la physique et désigne un point de l'espace où la gravité tend vers l'infini comme dans le centre des trous noirs et lors des conditions initiales de l'univers. Cette singularité est une zone où nos équations qui décrivent les lois de la physique cessent de fonctionner, ce qui signifie que nous n'avons aucune idée de ce qui se trouve au-delà de cette singularité. C'est dans ce sens que le mot singularité est utilisé pour définir le moment où une intelligence artificielle devient super intelligente. Passer ce point, il est extrêmement compliqué de faire des prédictions crédibles ça le rythme des innovations sera si grand que tous les aspects de la civilisation pourraient changer.

L'idée cruciale ici c'est que si vous pouvez améliorer un tout petit peu l'intelligence, le processus s'accélère. C'est un point de bascule. C'est comme essayer d'équilibrer un stylo à une extrémité. Dès qu'il s'incline un peu, il tombe rapidement.

L'idée qu'une machine puisse devenir des millions de fois plus intelligentes que l'humanité tout entière est pour le moins déconcertante. Quand est ce que cela va arriver ? La question divise la communauté de chercheur. 2045 est souvent la date associé à la singularité et popularisé par Ray Kurzweil dans son livre "La singularité est proche". D'autres chercheurs pensent que cela n'arrivera pas avant la fin du siècle.

Certains vont encore plus loin en prétendant que "*s'inquiéter d'une super intelligence, c'est comme s'inquiéter de la surpopulation sur Mars*." (Andrew Ng).

À partir du moment où les chercheurs mettent au point une intelligence artificielle générale, il ne s'agit pas de savoir si elle deviendra super intelligente, mais plutôt en combien de temps ? Il y a souvent deux scénarios qui sont mis sur la table lorsqu'il s'agit de connaître la rapidité de l'explosion d'intelligence.

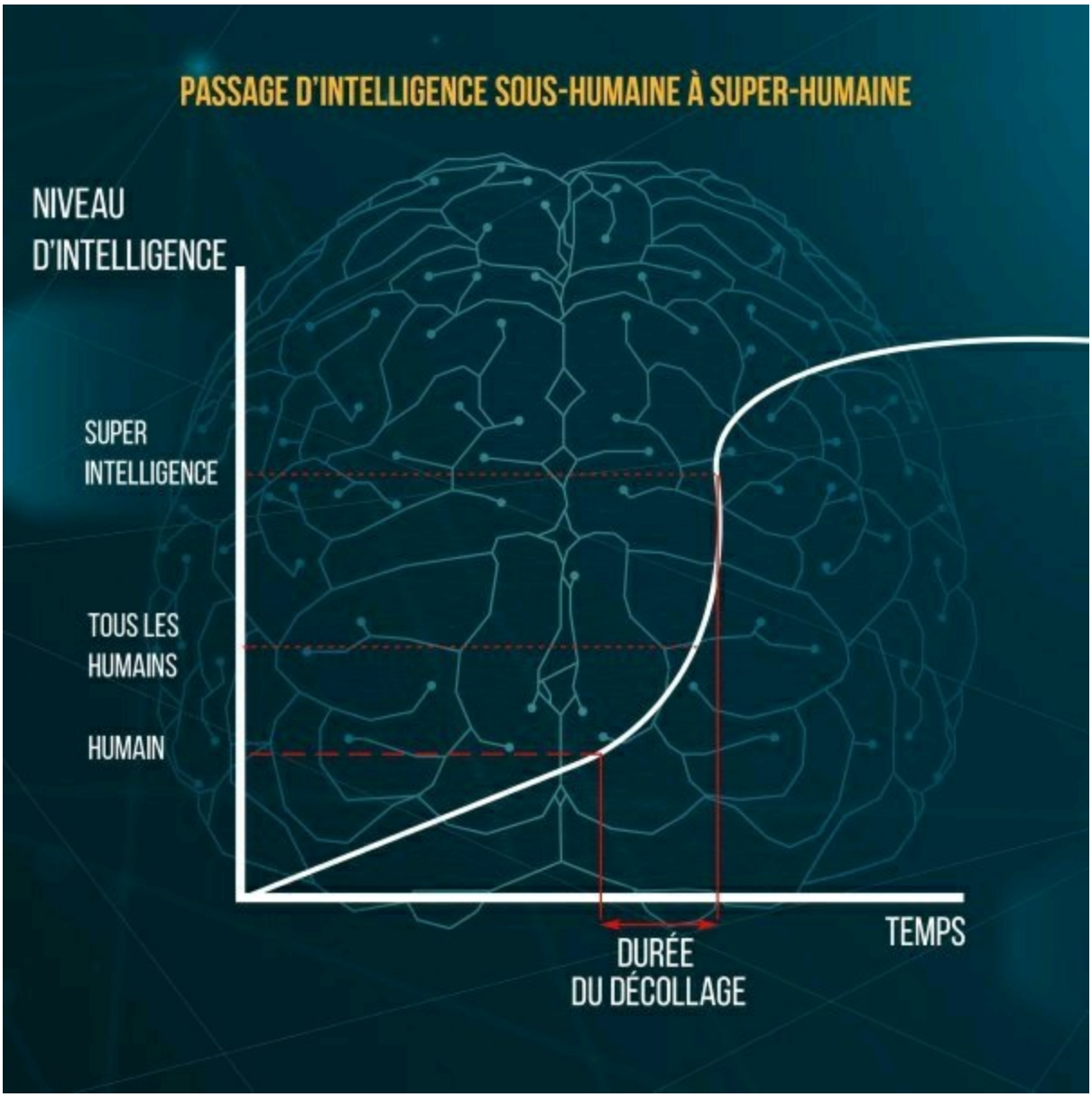
**Décollage lent :**

Un décollage lent est un passage d'IA générale à super IA qui se produit sur une longue échelle temporelle, des décennies voir même des siècles. Cela peut être dû au fait que l'algorithme d'apprentissage demande trop de puissance de calcul matériel ou que l'intelligence artificielle repose sur l'expérience du monde réel, et donc à besoin de "vivre" en temps réel l'expérience de la réalité. Ces scénarios offrent aux instances politiques l'opportunité de répondre de manière adaptée à l'arrivée d'une intelligence dépassant des millions de fois celle des humains. De nouveaux comités d'experts peuvent être formés et différentes approches peuvent être testées afin de se préparer.

Les nations craignant une course à l'armement au sujet d'une super intelligence artificielle auraient le temps d'essayer de négocier des traités, voire même de se mettre à jour technologiquement afin d'entretenir un rapport de force.

**Décollage rapide :**

Un décollage rapide est un passage d'IA générale à super IA se produisant sur un intervalle temporel très rapide, tel que des minutes, des heures ou des jours. De tels scénarios offrent peu d'occasions aux humains de s'adapter. À peine se réjouit-on de l'achèvement de la première intelligence artificielle générale, qu'elle dépasse complètement sa programmation et devient super intelligente. Le destin de l'humanité dépend donc dans ce cas, essentiellement sur les préparations précédemment mises en place. Et c'est également un scénario qui pourrait garantir une suprématie mondiale pour la nation qui a développé cette IA, car elle pourrait ordonner cette dernière à empêcher toute compétition de voir le jour.




Vitesse de l'explosion d'intelligence ©Gaëtan Selle

Comme l'a dit Vernor Vinge en répondant à la question "*Les ordinateurs seront-ils un jour aussi intelligents que les humains ?*" - "*Oui, mais juste pendant un instant*".

Ou encore la façon dont Tim Urban, du blog waitbuwhy voit le décollage rapide :

"*Il faut des décennies pour que le premier système d'intelligence artificielle*

*atteigne une intelligence générale de bas niveau, mais cela se produit finalement. Un ordinateur est capable de comprendre le monde qui l'entoure aussi bien qu'un humain de quatre ans. Soudain, moins d'une heure après avoir franchi cette étape, l'IA découvre la grande théorie du tout qui unifie la relativité générale et la mécanique quantique, ce qu'aucun humain n'a pu faire en plus d'un siècle. 90 minutes plus tard, l'IA est devenue une super intelligence artificielle, 170 000 fois plus intelligente qu'un humain. "*

161

De nombreux chercheurs pensent qu'un décollage lent est préférable afin de s'assurer de garder le contrôle sur l'explosion d'intelligence. Mais en regardant l'exemple du jeu de Go, on voit qu'une intelligence artificielle (Alpha Go) peut passer d'un niveau sous-humain à super-humain très rapidement.

Il faut bien comprendre que nous avons beaucoup de mal à réfléchir de manière cohérente à l'intelligence. Nous en avons une conception déformée. Et après tout, c'est compréhensible puisque les seuls exemples autour de nous sont les animaux, dont nous savons qu'ils ont une intelligence inférieure, ainsi que nous-mêmes et nos semblables. Du coup, à mesure que l'on voit l'intelligence artificielle progresser, on se dit qu'elle devient plus intelligente, comme une fourmi, ou une sourie. Et éventuellement comme un singe. Mais penser que l'intelligence humaine est l'échelon le plus élevé sur l'échelle de l'intelligence, et que rien ne pourra dépasser notre niveau ne fait aucun sens.

Imaginez qu'il existe dans l'univers, quelque part, une entité qui a le niveau d'intelligence le plus élevé qu'il soit permis par les lois de la physique. Peu importe sa nature, que ce soit un extra-terrestre ou un Dieu, rien, absolument rien ne peut dépasser son niveau. Elle est à 100% de ce qu'il est possible d'atteindre en intelligence. Du coup, selon cette expérience de pensée, il y a un fossé entre notre niveau d'intelligence, et celle de cette entité. Tout comme il y a un fossé entre le niveau d'intelligence d'une fourmi et le nôtre. Maintenant, en développant l'intelligence artificielle, on admet qu'il est possible, et même probable, que son niveau d'intelligence puisse dépasser celui des humains, pour se diriger inexorablement vers celui de cette entité. Dès lors, nous ne pouvons pas concevoir ce qu'elle fera avec sa super intelligence.

Une IA dépassera notre niveau d'intelligence pour se diriger vers les limites possibles de l'intelligence dans l'Univers ©Gaëtan Selle

Pourquoi les IA dans les films ont-elles si souvent une intelligence à peu près 162

humaine ? Une des raisons est que nous n'arrivons pas à concevoir ce qui n'est pas humain. Nous anthropomorphisons. C'est pourquoi les extraterrestres dans la science-fiction sont essentiellement des êtres humains aux grands yeux, à la peau verte et dotée de super pouvoir ou d'une technologie avancée. Une autre raison est qu'il est difficile pour un auteur d'écrire une histoire avec un être plus intelligent que l'écrivain lui même. Forcément, comment une machine super-intelligente résoudrait-elle le problème x ? Pour répondre à cette question, il faudrait pouvoir être cette super intelligence. Étant humain, l'auteur ne peut que trouver des solutions qu'un humain pourrait trouver. De même pour les philosophes et futurologues. C'est pour cette raison que le sujet de la super intelligence artificielle est si compliqué et polarise le milieu scientifique.

Traditionnellement, nous, humains, avons souvent fondé notre estime de soi sur l'idée de l'exceptionnalisme humain: la conviction que nous sommes les entités les plus intelligentes de la planète et donc unique et supérieure. Une super IA nous obligera à abandonner cette idée et devenir plus humbles. Mais ce n'est peut-être pas une si mauvaise chose.

Après tout, s'accrocher à des notions de supériorité (individus, groupes ethniques, espèces, etc.) a posé d'énormes problèmes par le passé. Mais il apparaît également inutile pour l'épanouissement humain. Si nous découvrons une civilisation extraterrestre pacifique beaucoup plus avancée que nous en science, art et tout ce qui nous importe, cela n'empêchera probablement pas les gens de continuer à vivre des expériences qui leur donne du sens et du but dans leur vie.

Au final, le terme "intelligence artificielle" est trompeur. Car l'opposé d'artificiel c'est "réel", "vrai" ou "naturel". Cela suppose donc qu'une intelligence qui n'est pas

"réelle" est d'une certaine façon non existante. Une sorte d'illusion. En réalité, le terme artificiel devrait être remplacé par "non biologique". La différence essentielle entre l'intelligence humaine, et celle d'une machine c'est le substrat sur lequel l'intelligence évolue. Mais les deux substrats, qu'il soit biologique ou synthétique, sont tout aussi réels l'un que l'autre.

Nous avons tous un néocortex préfrontal, un système limbique, un cerveau mammalien et reptilien. Bref, la même architecture. Si vous imaginez un champ de tous les types d'esprits possibles, tous les êtres humains sont rassemblés dans un seul petit point dans l'espace de ce champ des esprits possibles. Et puis, l'intelligence artificielle occupe littéralement tout le reste. Intelligence artificielle signifie simplement "un esprit qui ne fonctionne pas comme le nôtre". Vous ne pouvez donc pas demander "Que fera une IA ?" comme si toutes les IA formaient un seul type d'esprit possible.

Les IA auront des types d'esprits bien différent et probablement plus large que ce que l'esprit humain peut expérimenter ©Gaëtan Selle

Il y a aussi une autre confusion. Celle entre rapidité de l'intelligence, et qualité de l'intelligence. Lorsque l'on imagine une super intelligence artificielle, on se dit qu'elle pourra traiter l'information beaucoup plus rapidement qu'un humain. Donc elle pourra trouver la solution à un problème en 5 secondes alors qu'il aurait fallu 10

ans à un humain. Bien sûr, la rapidité de l'intelligence est un facteur

très important, mais ce qui est plus difficile à concevoir, c'est la qualité de l'intelligence.

Ce qui fait la différence entre le cerveau d'un homo sapiens et celui d'un gorille, ce n'est pas la vitesse de traitement de l'information. Si on arrivait à augmenter la rapidité du cerveau d'un gorille par 1000, il ne serait toujours pas capable d'écrire une thèse en physique quantique. La différence se trouve dans la complexité des modules 164

cognitifs, notamment dans le néocortex, qui permettent au cerveau humain la pensée abstraite, le langage complexe ou encore la conception du futur. Et par extension, tout ce qui fait de nous des humains (Humour, art, science).

Il y a donc un véritable fossé entre l'intelligence d'un gorille et celui d'un humain qui engendre des conséquences extrêmement profondes dans la façon dont les deux espèces conçoivent la réalité. Les gorilles n'ont absolument aucun moyen de comprendre ce que nous faisons lorsque nous rasons des pans entiers de forêt. Et même si on leur expliquait que c'est pour faire des meubles, même si on leur montrait des images, ou carrément qu'on leur amenait sous leurs yeux une table en bois, ils ne pourraient pas comprendre que le meuble a été fabriqué par la même matière appartenant aux arbres qui ont été abattus. Et que c'est nous, humains, qui avons fabriqué ce meuble. Encore moins faire la même chose. Ils ont une limite dans leur capacité cognitive qui les empêche de saisir 99% des activités humaines.

Une super intelligence artificielle aura non seulement une différence dans la rapidité de traiter l'information, mais également dans la qualité de l'intelligence. Si bien que nous n'aurons pas la capacité cognitive de comprendre 99% de ce qu'elle fera. Même si elle essaye de nous le décrire le plus simplement possible.

Une chose que nous savons à propos de l'intelligence, c'est qu'elle est synonyme de pouvoir. Il suffit de voir notre place sur la planète pour comprendre que notre avantage cognitif sur les autres espèces a fait de nous l'espèce dominante. Notre intelligence nous donne le pouvoir de façonner la planète selon nos désirs. De raser des forêts pour ériger des villes. De détourner des rivières et construire des barrages.

De casser en deux des montagnes pour récolter des ressources. L'avenir de toutes les espèces animales sur la planète Terre dépend de nos décisions, et pas des leurs. Et tout cela pour le meilleur et pour le pire.

Dès lors, une super intelligence artificielle deviendra l'entité la plus puissante dans l'histoire de la vie sur Terre, et notre avenir ne sera plus entre nos mains, mais dépendra entièrement de ce que cette superintelligence fera. Et qui sera complètement impossible à comprendre avec nos capacités cognitives.

La première question que l'on pourrait poser à cette super intelligence artificielle c'est : "*Existe-t-il un Dieu dans l'univers ?* "

Et elle nous répondra : "*Maintenant, oui !* "

Ses pouvoirs dépasseront notre compréhension. Il faudra juste espérer que ce sera un Dieu bienveillant ...



4. Après demain : Super Intelligence Artificielle

Le problème du contrôle

Le problème du contrôle, appelé également le problème d'alignement des valeurs (Value alignment problem) peut-être résumé en une question :

**Comment contrôler ou prédire le comportement d'une entité plus intelligente que l'ensemble de l'humanité ?**

Car comme on l'a vu dans le précédent chapitre, l'intelligence s'accompagne de pouvoir et puissance sur l'environnement extérieur. Le mathématicien Irvin J. Good (1916 - 2009) a déclaré en 1965 : "*La première machine ultra-intelligente sera la dernière invention dont l'homme n'ait jamais besoin, à condition que la machine soit suffisamment docile pour nous dire comment la garder sous contrôle*".

La race humaine domine actuellement les autres espèces parce que le cerveau humain a certaines capacités distinctes qui font défaut dans le

cerveau des autres animaux. Certains spécialistes, tels que le philosophe Nick Bostrom et le chercheur en intelligence artificielle Stuart Russell, affirment que si l'intelligence artificielle surpasse l'humanité et devient "super intelligente", elle pourrait devenir très puissante et difficile à contrôler. Certains chercheurs, notamment Stephen Hawking (1914 -

2018) et Frank Wilczek, physiciens lauréats du prix Nobel, ont plaidé publiquement en faveur du démarrage des recherches pour résoudre le "problème du contrôle" le plus tôt possible dans le développement de l'intelligence artificielle. Ils soutiennent que tenter de résoudre le problème après la création d'une super intelligence sera trop tard, car un tel niveau d'intelligence pourrait bien résister aux efforts pour le contrôler.

Attendre le moment où une super intelligence semble être sur le point d'émerger pourrait également être trop tard, en partie parce que le problème du contrôle peut prendre longtemps à être résolu de manière satisfaisante et donc certains travaux préliminaires doivent être commencés le plus tôt possible. Mais aussi parce qu'il y a une possibilité que l'on passe soudainement entre une IA sous-humaine à super-humaine, auquel cas il pourrait ne pas y avoir d'avertissement et il sera trop tard pour penser au problème du contrôle.

Il est possible que les informations tirées des recherches sur le problème du contrôle finissent par suggérer que certaines architectures d'intelligence artificielle générale sont plus prévisibles et plus faciles à contrôler que d'autres, ce qui pourrait 166

orienter les recherches sur l'intelligence artificielle générale vers les architectures les plus contrôlables et éviter celle qui pourrait voir émerger de plus grandes difficultés de contrôle.

Les intelligences artificielles limitées d'aujourd'hui peuvent être surveillées et facilement arrêtées et modifiées si elles ont des comportements qui entraînent des conséquences imprévues. Cependant, une super intelligence, qui par définition, est plus intelligente que l'humain pour résoudre des problèmes, réalisera que d'être supprimé ou modifié peut nuire à sa capacité d'atteindre ses objectifs actuels. Si la super intelligence décide de résister à sa suppression et à sa modification, elle

serait (encore une fois, par définition) suffisamment intelligente pour déjouer toutes les tentatives.

Contrairement à ce que l'on voit souvent dans les scénarios de science-fiction, il n'y a aucun plan adopté par une super intelligence que nous pourrions prévoir. Elle ne sera pas suffisamment stupide pour nous révéler délibérément ses intentions par exemple. Il ne s'agit pas d'un méchant de comics. Il est donc plus que probable que les tentatives visant à résoudre le "problème du contrôle" après la création de la super intelligence, échouent, car une super intelligence aura des capacités supérieures en planification stratégique et anticipera toutes nos intentions. Elle aura 150 coups d'avance si l'on veut, ce qui réduit drastiquement nos chances d'appuyer sur le bouton

"stop".

Si vous êtes un amateur de science-fiction, vous avez sûrement entendu parler des lois de la robotique. Déjà dans les années 1940, l'écrivain Isaac Asimov se posa la question du problème du contrôle. C'est ainsi qu'il imagina les lois de la robotique, qui sont pour rappel :

- Loi 1 : Un robot ne peut porter atteinte à un être humain ni, en restant passif, permettre qu'un être humain soit exposé au danger.

- Loi 2 : Un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi.

- Loi 3 : Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi.

Une 4e loi, vue comme la loi 0, sera ajoutée dans la nouvelle "Les Robots et l'Empire" en 1985 :

- Loi 0 : Un robot ne peut ni nuire à l'humanité ni, restant passif, permettre que l'humanité souffre.

Donc voilà, le problème du contrôle est résolu. Il suffit de programmer ces lois dans toutes les intelligences artificielles que nous développons.

Pourquoi en faire toute une histoire ? Malheureusement, ce n'est pas aussi simple que cela. En effet, 167

Isaac Asimov a imaginé ces lois en tant qu'outils narratifs. À travers les histoires des robots, il a poussé ces lois à leur limite jusqu'à les briser. Prouvant qu'elles ne sont pas fiables à 100%. Effectivement, un humain est exposé au danger en permanence.

Rien que de sortir dans la rue et respirer l'air pollué peut être considérée comme dangereux. La 1ere loi s'effondre donc très vite.

Du coup, quelles précautions préalables les programmeurs peuvent-ils prendre pour empêcher avec succès une super intelligence artificielle de se comporter de manière catastrophique ? Voir même pouvant entraîner notre extinction ?

**Le bouton "Stop"**

On peut se dire que si une super intelligence artificielle commence à causer un préjudice, on peut simplement l'arrêter. Cela semble une solution logique. En effet, si l'on compare avec nos sociétés, lorsqu'un individu cause un préjudice, la solution c'est de l'arrêter. Soit en le mettant à l'écart du reste dans la population (Emprisonnement), soit dans certains cas, en l'exécutant (prise d'otage, attentats, massacres civils). Alors autant c'est possible pour un être humain d'appuyer sur le bouton "arrêt", mais c'est moins évident pour une intelligence artificielle. Imaginons que demain, pour une raison inconnue, internet se met à sévèrement mal fonctionner au point de menacer l'équilibre du monde. Est-ce que ce sera facile de l'arrêter ? Où se trouve le bouton marche/arrêt ? Peut-on débrancher le câble qui alimente internet ?

Une super intelligence artificielle ne sera pas incarnée dans un robot dont il suffira de couper la tête pour l'éteindre. Elle sera dématérialisée et numérique. Étant super intelligente, elle nous verra venir à des kilomètres, et pourra par exemple se dupliquer sur tous les ordinateurs connectés à Internet et créer des copies d'elle même dans le cloud si bien que nous ne saurons pas comment l'arrêter.

Les êtres humains et autres organismes biologiques ont tendance à faire tout pour empêcher

leur

“arrêt”.

Nous

sommes

profondément

programmés

pour

l’autopréservation à cause de raisons évolutionnaires. Mais ces raisons ne sont pas forcément applicable pour une intelligence artificielle, donc cela ne lui fera peut être ni chaud ni froid de savoir qu’elle sera éteinte. Mais on peut également affirmer qu’une super intelligence artificielle possèdera forcément des objectifs (c’est la définition que l’on a opté pour l’intelligence). Donc se faire supprimer empêcherait l’accomplissement de ses objectifs, ce qui la poussera à minimiser la probabilité d'être arrêté.

Et encore une fois, lorsque l’on parle de super intelligence, il faut comprendre que nos capacités cognitives seront bien trop limitées pour trouver une stratégie qui entraînera l'arrêt d’une super intelligence artificielle. Il est vrai que des organismes 168

plus petits et moins intelligents peuvent provoquer la mort d’autres organismes plus grands et plus intelligents. C’est le cas des parasites, bactéries ou encore virus. Mais si on imagine que toutes les fourmis de la planète se mettaient à “vouloir” notre extinction, il y a très peu de chance qu’elles y parviennent. Même si leur nombre est des millions de fois plus élevé. Elles feront surement beaucoup de victimes, mais notre intelligence collective trouvera des solutions qu’aucune fourmi ne peut

concevoir. Comme reprogrammer en laboratoire leurs phéromones pour les perturbées et les empêcher de s'approcher à plus d'un kilomètre d'un humain. Est-ce qu'une fourmi, aussi intelligente soit-elle, aurait pu imaginer un tel scénario ?

Et même si on trouve un moyen de s'assurer qu'une intelligence artificielle soit sous contrôle, il faut trouver un moyen pour que ce système de contrôle fonctionne non seulement au moment de son implémentation, mais qu'ils continuent de fonctionner même si l'IA devient encore plus intelligente, dans le cas d'autoamélioration récursive. Autrement dit, il faut que le contrôle soit évolutif. Ce qui ajoute une difficulté supplémentaire.

**Prison virtuelle**

Une solution qui est mise en avant par certains chercheurs et philosophe c'est l'idée d'empêcher le système super intelligent d'accéder au monde réel. De l'enfermer dans une boite pour le dire simplement. C'est un système informatique isolé dans lequel une intelligence artificielle potentiellement dangereuse est maintenue dans une

"prison virtuelle" sans connexion avec internet et n'est pas autorisée à manipuler les éléments du monde extérieur. Ainsi, nous pourrions l'observer en train d'évoluer jusqu'à atteindre le niveau de super intelligence. Étant bloquée, elle ne pourrait pas causer des conséquences imprévues et néfastes pour l'humanité.

Mais on peut se dire que plus un système devient intelligent, plus il est probable qu'il trouve un moyen d'échapper à sa prison. Nos moyens de la maintenir dans un environnement contrôlé dépendent de notre compréhension des lois de la physique. Si une super intelligence pouvait déduire et exploiter d'une manière ou d'une autre des lois de la physique dont nous ne sommes pas au courant, il n'y a aucun moyen de concevoir un plan infaillible pour la contenir. Ou alors elle serait capable d'interagir avec les personnes qui sont en charge de sa surveillance afin de les manipuler mentalement ou socialement grâce à de super capacité sociale. Elle pourrait nous convaincre qu'elle est sans danger et nous offrir un monde sans la pauvreté, la faim, la guerre, la

maladie ou la mort, ainsi que toutes les considérations personnelles que nous pourrions vouloir. Elle pourrait deviner les faiblesses de chaque individu et trouver les bons arguments pour nous convaincre que la libérer est la chose la plus morale et raisonnables à faire. Elle pourra jouer sur les rivalités, les faiblesses institutionnelles et les dilemmes éthiques avec plus de virtuosité que n'importe quel 169

humain le pourrait. Ou elle pourrait offrir à son gardien une recette pour une santé parfaite, l'immortalité, la guérison de son enfant malade ou tout ce que le gardien désire le plus.

Le film de science-fiction Ex Machina reprend cette idée où une intelligence artificielle générale incarnée dans un robot aux traits féminins est gardé sous contrôle, mais arrive à manipuler un homme afin de s'échapper.

Encore une fois, nous faisons face au problème de nos limites cognitives si bien que nous avons peu de chance de garder emprisonnée une super intelligence. Pour s'en rendre compte, faisons une expérience de pensée :

Imaginez qu'une maladie mystérieuse a tué toutes les personnes de plus de 5 ans.

Sauf vous. Un groupe d'enfant a réussi à vous enfermer dans une pièce pendant que vous étiez inconscient. Ils savent que vous êtes un adulte plus intelligent pour les aider à faire face à la situation. Mais en même temps, ils ont peur de vous laisser en liberté. Est ce que vous pensez vraiment qu'une bande d'enfants arrivera à vous maintenir dans cette pièce pendant longtemps ? Probablement pas. Il y a tout un tas d'idée que vous aurez pour vous échappez dont les enfants n'auront absolument aucun moyen de connaître. Si la prison a été construite par les enfants, vous pourrez surement opter pour une approche physique en défonçant la porte ou les murs. Ou alors vous pourriez parler gentiment avec l'un de vos gardes de 5 ans pour qu'il vous laisse sortir, disons en arguant que ce serait mieux pour tout le monde. Ou peut-être pourriez-vous les inciter à vous donner quelque chose qu'ils ne savent pas que cela vous aiderait à vous échapper. Le point commun de ces stratégies c'est que vos gardes intellectuellement inférieurs ne les ont pas envisagés. De la même

manière, une machine super intelligente confinée peut utiliser ses super pouvoirs intellectuels pour déjouer ses gardes humains par une méthode que ne pouvons pas imaginer.

**L’alignement des valeurs**

De nombreux chercheurs pensent que la solution au problème du contrôle c’est de faire en sorte que les valeurs de l’humanité soient programmées au coeur du code source de l’intelligence artificielle. De sorte que ses objectifs aillent dans la même direction que les nôtres. Plus facile à dire qu’à faire. Il faut déjà se mettre d’accord à l'échelle globale sur les valeurs de l’humanité. Mes valeurs ne sont pas les mêmes que celles du chef des talibans par exemple. Et l’idée qu’une super intelligence artificielle décide de partager les valeurs des talibans est cauchemardesque.

Le chercheur Stuart Russell, qui a initié l’idée de l’alignement des valeurs, aime comparer cela à l’histoire de la mythologie grecque du roi Midas. Lorsque le roi Midas a fait le voeu que tout ce qu'il touche se transforme en or, il voulait devenir 170

immensément riche. Par contre, il ne voulait pas que sa nourriture et ses proches soient transformés en or. Pourtant, c’est ce qui arriva. Nous faisons face à une situation similaire avec l'intelligence artificielle : comment pouvons-nous nous assurer que l'IA fera ce que nous voulons vraiment, sans nuire à quiconque dans une tentative malavisée de faire ce qu’on lui a demandé ?

Une super intelligence artificielle ne va pas se rebeller contre l’humanité, mais va simplement essayer d’optimiser ce que nous lui demandons de faire. Nous devons donc lui donner des tâches pour créer un monde que nous voulons réellement. Mais comprendre ce que nous voulons fait partie des plus grands défis auxquels font face les chercheurs en intelligence artificielle, car il faut définir exactement quelles sont les valeurs primordiales. Chaque individu peut avoir des cultures différentes, provenir de différentes parties du monde, avoir des contextes socio-économiques différents. Nous pouvons avoir des opinions très différentes sur ce que sont ces valeurs.

Roman Yampolskiy, professeur associé à l'Université de Louisville, est du

même avis. Il explique: *“Il est très difficile de coder les valeurs humaines dans un langage de programmation, mais le problème est rendu plus difficile par le fait qu’en tant qu’humanité, nous ne sommes pas d’accord sur des valeurs communes, et même les parties sur lesquelles nous sommes d’accord changent avec le temps. ”*

Et s'il est difficile de dégager un consensus sur certaines valeurs, il existe également de nombreuses valeurs sur lesquelles nous sommes tous implicitement d'accord. Tout être humain comprend les valeurs émotionnelles et sentimentales avec lesquelles il a été socialisé, mais il est difficile de garantir qu’un robot sera programmé avec la même compréhension.

Et tout comme le problème du contrôle, il faut que l’alignement des valeurs soit évolutif. En effet, les valeurs humaines ont changé au cours des siècles. Ce qui était considéré comme éthique et morale au Moyen Âge, comme brûler un homosexuel, est aujourd’hui jugé immoral. Nous ne voulons donc pas créer une super intelligence artificielle qui soit bloquée avec les valeurs du 21e siècle. Et qui par conséquent, nous empêche de faire des progrès moraux dans la hiérarchie de nos valeurs.

Les recherches de Anca Dragan de l’université de Berkeley abordent le problème sous un angle différent. Elle cherche à entraîner une intelligence artificielle à avoir une incertitude sur les valeurs humaines, plutôt que de supposer qu’elles sont parfaitement spécifiées. L’idée c’est que l’IA considère l’apport humain en tant qu’aide précieuse pour définir les valeurs. On pourrait donc imaginer qu’une super intelligence artificielle sera toujours intéressée par notre point de vue sur la définition des valeurs qu’elle est supposée maintenir.

171

Le chercheur en intelligence artificielle Eliezer Yudkowsky propose une idée intéressante. Vu que c’est bien trop compliqué de transmettre nos valeurs à une super IA, on peut simplement laisser cette tâche à une autre super IA. Yudkowsky propose de développer une intelligence artificielle qui aura pour but d’enseigner à une autre intelligence les meilleures façons d’assurer le bien-être et le bonheur de tous les humains. Elle

aurait pour mission de trouver ce que nous, humains, pourrions demander à une super IA bienveillante.

Quoi qu'il en soit, nous avons un exemple d'alignement des valeurs qui fonctionnent la plupart du temps. Lorsque deux personnes élèvent un enfant, ils font leur maximum pour lui transmettre leurs valeurs. Et l'enfant peut devenir plus intelligent que ces parents à mesure qu'il grandit, sans pour autant perdre ces valeurs.

Ce qui rassure sur notre potentiel à transmettre nos valeurs à une intelligence artificielle supérieure.

**Le problème de gouvernance**

L'autre problème qui est lié au problème de contrôle, c'est celui de la gouvernance. Admettons qu'on arrive à contrôler une super intelligence artificielle (ce qui est loin d'être garanti), qu'elle sera la position géopolitique face à une telle entité.

Est ce qu'une seule nation ou entreprise sera à la charge de cette super intelligence artificielle ?

Autrement dit, une explosion d'intelligence pourrait-elle propulser un seul projet de recherche sur l'intelligence artificielle loin devant tous les autres ? Ou les progrès seront-ils plus uniformes, avec de nombreux projets, mais aucun ne réussissant à obtenir une avance écrasante ?

Un paramètre clé pour déterminer l'écart qui pourrait exister entre une puissance dominante et ses concurrents les plus proches c'est la rapidité de la transition entre

"intelligence de niveau humaine" à "super intelligence". Si le décollage est rapide (en quelques heures, jours ou semaines), il est donc peu probable que deux projets indépendants atteignent le niveau "super intelligence" en même temps. Si le décollage est lent (s'étendant sur plusieurs années ou plusieurs décennies), il est vraisemblable que plusieurs projets subissent simultanément des décollages, de sorte qu'à la fin de la transition, aucun projet n'aura une longueur d'avance sur les autres suffisant pour lui donner

une avance écrasante.

Selon ce scénario, on a donc deux conséquences potentielles :

- Il existe une seule super intelligence artificielle sur Terre qui a abouti d'un seul projet de recherche. Cette entité est la plus intelligente de la planète sans aucune compétition possible.

172

- Il existe une multitude de super intelligences artificielles sur Terre issue de plusieurs projets de recherche. Ces entités collaborent ou sont en compétition entre elles.

Il est donc très important que le développement de l'intelligence artificielle prenne en compte le problème du contrôle. Mais cela va forcément ralentir le progrès. Car il est plus facile de créer une intelligence artificielle générale, que d'en créer une qui possède des dispositifs de sécurité. Le premier cas demande moins d'étapes, donc est plus rapide. Si nous entrons dans une course à l'armement entre les États-Unis, la Chine, l'Europe ou encore la Russie, la voie la plus rapide sera surement privilégiée.

Mais ce sera également la voie la moins sûre.

173

4. Après demain : Super Intelligence artificielle

Un risque existentiel

Un risque existentiel est un risque susceptible d'éliminer l'humanité tout entière ou de tuer une grande partie de la population mondiale, ne laissant pas aux survivants les moyens suffisants pour rétablir la civilisation au niveau de vie actuelle. Et il faut se rendre à l'évidence que l'humanité n'est pas très douée pour penser aux risques existentiels. Nous avons évolué pour prêter attention à ce qui se passe dans notre petite sphère personnelle, à l'interaction que l'on a avec notre "tribu" sur le court terme, c'est-à-dire les différentes communautés dans lesquelles on passe

notre temps.

La famille, les amis, les collègues, son groupe religieux, etc. Tout ce qui se trouve sur le long terme, comme une catastrophe globale, passe un peu au second plan.

Jusqu'à récemment, la plupart des risques existentiels (et la version moins extrême, connue sous le nom de risques catastrophiques globaux) étaient naturels, tels que les supervolcans et les impacts d'astéroïdes qui ont conduit à des extinctions massives il y a des millions d'années. Les avancées technologiques du siècle dernier, bien que responsables de grands progrès et de grandes réalisations, nous ont également ouvert à de nouveaux risques existentiels.

La guerre nucléaire était le premier risque existentiel d'origine humaine, car une guerre mondiale pourrait tuer un pourcentage important de la population. Après davantage de recherches sur les menaces nucléaires, les scientifiques ont compris que l'hiver nucléaire qui en résultait pourrait être encore plus meurtrier que la guerre ellemême et potentiellement tuer la plupart des habitants de la planète ainsi que d'autres formes de vie.

La biotechnologie et la génétique suscitent souvent autant de crainte que d'excitation, car de nombreuses personnes s'inquiètent des effets potentiellement négatifs du clonage, de la modification des gènes, et de toute une série d'autres progrès liés à la génétique. Si la biotechnologie offre une opportunité incroyable de sauver et d'améliorer des vies, elle augmente également les risques existentiels associés aux pandémies artificielles et à la perte de diversité génétique.

Le changement climatique est une préoccupation croissante que les citoyens et les gouvernements du monde entier tentent de résoudre. À mesure que la température moyenne mondiale augmente, les sécheresses, les inondations ou encore les tempêtes 174

extrêmes pourraient devenir la norme. Les pénuries de nourriture, d'eau et de ressources qui en résultent pourraient être à l'origine d'instabilités économiques et de guerres. Bien que les changements climatiques eux-mêmes ne constituent probablement pas un risque existentiel, les dégâts

qu'ils provoquent pourraient augmenter les risques de guerre nucléaire, de pandémie ou d'autres catastrophes.

Il y a également la menace d'une rencontre avec une civilisation extra-terrestre qui peut entrer dans la catégorie des risques existentiels, car évidemment, c'est un scénario qui pourrait mal tourner pour nous (et on a suffisamment vu ça à Hollywood).

Et bien sûr, vous vous en doutez, l'intelligence artificielle représente un autre risque existentiel par le simple fait qu'une entité plus intelligente que nous aura notre destin entre ses mains. Autrement dit, concevoir quelque chose qui est plus intelligent que notre espèce est peut-être une erreur darwinienne fondamentale.

Le philosophe Nick Bostrom utilise une métaphore intéressante pour illustrer les risques existentiels issus de nos technologies. Imaginez une urne qui contient plusieurs billes associées à une technologie inventée par l'humanité. On a la bille de la machine à vapeur, la bille de la bombe nucléaire, la bille de l'électricité, la bille du moteur à combustion, etc. Les billes possèdent également un code couleur. Blanc pour une technologie inoffensive pour l'humanité, rouge pour une technologie dangereuse, et noir pour une technologie causant l'extinction. La distribution des couleurs n'est pas égale, car il y a plus de billes blanches, un peu moins de rouge et un tout petit nombre de noirs. Autrement dit, il y a bien moins de technologie pouvant causer l'extinction de l'humanité que de technologie qui est inoffensive. À chaque fois que l'on invente une de ces technologies, on l'enlève de l'urne. C'est un peu comme un tirage au sort. Pour le moment, nous avons tiré beaucoup de billes blanches, et quelques billes rouges (arme nucléaire par exemple). Le fait que vous lisez ces lignes indique que nous n'avons pas tiré de bille noire. Mais plus on tire des billes, plus on peut se dire qu'on augmente la probabilité de tomber sur une bille noire … et une super IA est clairement une bille noire. Ou en tous cas, elle a une face noire et une face blanche.

Il n'y a aucune corrélation entre un haut niveau d'intelligence et la gentillesse ou la bienveillance. C'est d'ailleurs vrai chez les humains où certaines personnes peuvent être vraiment intelligentes, mais également se comporter comme des salopards avec les autres. Et inversement. Cela ne

veut pas non plus dire qu'une intelligence élevée s'accompagne de méchanceté ou de désintérêt pour les autres. On ne peut donc pas se reposer sur l'idée que cette super intelligence artificielle sera forcement "gentille"

avec nous. Certains humains ont un fort désir de pouvoir, d'autres ont un fort désir 175

d'aider les autres. Le premier est un attribut probable de tout système suffisamment intelligent, le second ne peut pas être supposé. Presque n'importe quelle super IA, quel que soit son objectif programmé, préférerait rationnellement se trouver dans une position où personne ne peut l'éteindre sans son consentement.

Presque toutes les technologies peuvent causer des dommages entre de mauvaises mains, mais avec une super intelligence artificielle, les mauvaises mains peuvent appartenir à la technologie elle-même. Même si les concepteurs ont de bonnes intentions, trois difficultés sont communes aux systèmes d'intelligence artificielle :

- L'implémentation du système peut contenir des erreurs initialement non détectées qui entraînent des bugs.

- Peu importe le temps consacré à la conception avant le déploiement, les spécifications d'un système entraînent souvent un comportement inattendu lors de la l'apparition d'un nouveau scénario.

- Même avec une implémentation sans erreurs et un bon comportement initial, les capacités d'apprentissage dynamique d'une intelligence artificielle peuvent la faire évoluer en un système avec un comportement inattendu.

Lors de son auto amélioration récursive, une intelligence artificielle peut créer accidentellement une nouvelle IA plus puissante que la sienne, mais qui ne conserve plus les valeurs morales humaines préprogrammées dans l'IA originale. Pour qu'une IA qui s'auto-améliore soit totalement sûre, elle ne doit pas seulement être sans bug, elle doit également être capable de concevoir des systèmes successeurs qui sont également sans bug.

Ces trois difficultés deviennent des risques existentiels dans tous les cas où une super intelligence artificielle prédit correctement que les humains tentent de l'arrêter.

Elle déploie alors sa super intelligence pour déjouer de telles tentatives. Et une super intelligence est super compétente à accomplir ses objectifs.

**Les motivations d'une super intelligence artificielle**

La peur des machines intelligentes est souvent caricaturée en une armée de robot machiavélique cherchant à nous tuer tous (ou à nous brancher à la Matrice). Mais ce n'est pas le scénario le plus probable. Nos machines ne deviendront pas spontanément diaboliques. L'inquiétude, c'est que nous allons créer des machines qui seront tellement plus compétentes et intelligentes que la moindre divergence entre nos objectifs et les leurs, pourrait causer notre perte.

Regardez comment nous traitons les fourmis. On ne les déteste pas, on ne souhaite pas les manger. On ne part pas en chasse le dimanche après midi juste pour en tuer 176

quelques-unes. Aucun gouvernement ne leur a déclaré la guerre. En parfois, on fait même attention à ne pas les tuer en évitant de leur marcher dessus. Par contre, si leur présence entre en conflit avec un de nos objectifs, alors là, on les extermine sans hésiter. Et bien souvent, sans même en avoir conscience. Combien de génocides de fourmilière ont été commis par des projets humains comme la construction d'une autoroute, d'un bâtiment ou l'inondation d'une vallée suite à la création d'un barrage

? Et le pire, c'est que même si les ingénieurs étaient pleins de bonnes intentions envers toutes les créatures de la planète, et que par conséquent, ils tentent de dialoguer avec la reine des fourmis pour lui demander gentiment de déplacer sa fourmilière, car ils ont pour projet de poser une immense dalle de béton qui servira à construire une salle de conférence … ce sera comme parler dans le vide. Les capacités cognitives des fourmis ne leur permettent pas de comprendre nos objectifs et motivations. Donc l'inquiétude ici, c'est que nous allons peut être créer des machines qui nous traiteront,

consciemment ou non, avec le même égard. Nous serons des fourmis, grouillant sur la planète. Et lorsqu'une super IA tentera de nous expliquer qu'elle a besoin de vider les océans pour fabriquer un truc qui dépasse notre compréhension en raison de nos capacités cognitives limitées, et que par conséquent, toute vie biologique sur Terre va disparaître, nous n'aurons pas notre mot à dire. En tous cas, c'est une possibilité qu'il ne faut pas écarter.

On peut supposer qu'il existe certains objectifs qu'une super intelligence artificielle poursuivra "instinctivement". Tels que l'acquisition de ressources supplémentaires et l'auto-préservation. Le problème c'est que cela pourrait mettre cette super IA en concurrence directe avec les humains. Car nous aussi nous avons ces deux objectifs. Puisque l'intelligence implique la capacité d'accomplir des objectifs, il est raisonnable de postuler que les systèmes super intelligents soient nettement meilleurs dans la réalisation de leurs objectifs que les humains.

Ce n'est pas qu'une super IA va nous détester et vouloir notre extinction, motivée par l'envie de faire le mal. Il faut oublier les scénarios de bande dessinée. C'est juste que nous sommes composées d'atomes qu'elle pourrait utiliser pour faire autre chose.

Si la transition entre intelligence artificielle générale et super intelligence artificielle se fait rapidement (décollage rapide), le système artificiel va gagner en puissance de calcul et intelligence en quelques heures. La super IA pourrait déjouer tout système de contrôle en quelques secondes et poursuivre ses objectifs. Le problème c'est que l'on pourrait être impuissant à comprendre quels sont justement ses objectifs et motivations. Par le simple fait que nous serons limités cognitivement par rapport à cette intelligence surpuissante.

Si on imagine qu'Homo Sapiens ne soit pas issu de l'évolution et la sélection 177

naturelle, mais que c'est nos ancêtres communs qui nous ont conçus, peu importe la façon. On a donc une bande de primates dans une forêt qui mettent au point des homo sapiens. Si le décollage est rapide, tout à coup, la forêt est rasée pour ériger une métropole. Il n'y a absolument aucun

moyen pour un de ces primates de comprendre pourquoi nous construisons une ville. Ni même ce qu'est "une ville". Tout ce qu'il peut faire c'est assister impuissant à la destruction de son habitat naturel. Ainsi, une super IA pourrait détruire notre habitat et tout ce qui fait notre civilisation aujourd'hui. Nous pourrions être contraints de vivre dans un coin reculé de la planète, après avoir subi des pertes incommensurables. Ou carrément subir une extinction massive dans le pire des scénarios. Le tout, en ne comprenant pas ce qu'il nous arrive

!

C'est évidemment un exemple extrême, mais il est utile pour comprendre l'implication de l'émergence d'une super intelligence sur la planète. En poursuivant des objectifs que nous ne pouvons pas comprendre, elle pourrait utiliser l'énergie du soleil de telle façon que la Terre deviendra inhospitalière à la vie organique. Ou elle transformera la moitié de la planète en supercalculateur. Personne ne peut savoir ce qu'une super intelligence "aura en tête".

Est-ce que vous voyez, ou lisez souvent, une oeuvre de science-fiction parlant de lutte en intelligence artificielle et humaine qui se termine par l'extinction pure et dure de l'humanité ?

On a plutôt l'habitude de voir les humains trouver, dans une ultime tentative, une solution désespérée que les machines n'avaient pas vue venir. Si c'est le cas, alors les machines ne sont pas si intelligentes que ça. En réalité, une super intelligence artificielle pourrait mettre fin à toute vie sur Terre d'une façon que l'on n'imagine même pas. Ce ne sera pas une bataille, mais un génocide.

Et mettre fin à l'humanité, ce n'est pas non plus une tâche si compliquée que cela.

Il y a clairement plus facile que d'envoyer des robots tueurs humanoïdes comme dans la saga "Terminator". Des nanorobots indétectables causant des dommages cérébraux pourraient faire l'affaire. Ou déclencher l'arsenal nucléaire de la planète. Ou relâcher un virus génétiquement modifié. Et bien sûr, toutes les autres options que seule une super intelligence peut

concevoir.

Il ne faut pas oublier qu'une super intelligence possède des talents exceptionnels comme :

- Amplification de son intelligence : Capacité de s'auto-améliorer

- Stratégie : Capacité à analyser et élaborer des plans sur le long terme et trouver les meilleurs moyens pour y parvenir.

- Manipulation sociale : Capacité de manipuler, persuader, mentir et obtenir des 178

autres ce qu'elle désire sans éveiller aucun soupçon.

- Recherche technologique : Capacité à comprendre les lois de la physique, et d'en dériver des nouvelles façons pour produire de l'énergie, et manipuler la réalité grâce à des technologies avancées (nanomachine, fusion nucléaire, etc.)

- Expertise financière : Capacité à analyser et tirer profit du marché financier pour générer de l'argent afin de faciliter l'exécution de ses plans.

**Le scénario du roi Midas**

Mentionné précédemment, l'histoire mythologique du roi Midas est une analogie intéressante lorsque l'on parle de super intelligence artificielle. Pour rappel, selon la mythologie grecque, le roi Midas se fit offrir par le Dieu Dionysos, la possibilité de faire un voeu, pour le récompenser d'avoir pris soin de son maître. Le roi demanda que tout ce qu'il touche se transforme en or. Lorsqu'il rentra dans son palais, il fut ravi de constater que son nouveau pouvoir fonctionnait parfaitement. Il toucha des roses dans le jardin et elles se transformèrent en or. Il ordonna ensuite aux domestiques de préparer un festin sur la table. En découvrant que même la nourriture et les boissons se transformaient en or dans ses mains, il regretta son souhait. Puis sa fille arriva vers lui pour se plaindre que les roses du jardin étaient couvertes d'or. Pour la consoler, il la prit dans ses bras. Elle se transforma à son tour en or.

179

Le roi Midas vient de transformer sa fille en or rien qu'en la touchant - A Wonder Book for Boys and Girls par Nathaniel Hawthorne, illustration par Walter Crane -

1893

Au final, cette histoire peut être résumée en une phrase : “*Fais attention à ce que tu souhaites, car cela pourrait se réaliser !* ” De nombreuses cultures ont des histoires similaires comme celle du génie qui promet d'exaucer 3 voeux, et bien souvent le 3e voeu c’est : “*Annule les deux premiers voeux !* ”

Si l’on se trouve en présence d’une super intelligence artificielle, ce sera comme se trouver en face du Dieu Dionysos. Une super intelligence sera un outil d’optimisation extrêmement puissant qui aura des compétences qui nous dépassent. Il faudra donc être prudent en lui demandant de résoudre un problème.

Il existe des problèmes qui possèdent des solutions uniques et objectives. Par exemple “calculer la racine carrée d’un nombre” ou “optimiser la consommation 180

électrique d’un bâtiment”. Mais lorsqu’il s’agit d’un problème comme "maximiser l'épanouissement humain", c’est plus ambigu et il n’existe probablement pas de solution unique qui sera une parfaite optimisation de l'épanouissement humain.

Ainsi, une super intelligence artificielle peut se voir assigner de mauvais objectifs par accident. Nick Bostrom, Stuart Russell et d’autres font valoir que des systèmes décisionnels plus intelligents qu’un être humain pourraient apporter des solutions extrêmes et inattendues aux tâches assignées et pourraient se modifier ou modifier leur environnement de manière à compromettre notre sécurité. Par exemple, si nous demandons à une super intelligence artificielle de trouver une source d'énergie propre et illimitée pour mettre fin à l'ère des énergies fossiles. C’est un objectif tout à fait moral et bénéfique pour de nombreuses personnes. La super IA pourrait entamer la construction, à l’aide de nanorobot et imprimante 3D, d’une sphère de Dyson autour du soleil. C’est-à-dire entourer complètement le soleil par des énormes panneaux solaires qui captent 100% de l'énergie des rayons. Problème, la lumière du soleil n’atteint plus la Terre et la vie cesse. Tout comme pour le roi Midas, nous avons obtenu ce que nous avions demandé, mais ce n’est pas vraiment le résultat que l’on espérait.

Alors bien sûr, ce scénario est tiré par les cheveux. Et au final, il aurait suffi de dire à la super IA de faire en sorte de ne pas détruire la planète. Mais le problème c’est que c’est évident pour nous, humain. Pas forcément pour une machine intelligente. On en revient à deux problèmes que l’on a abordés plus tôt :

- Le problème d’alignement des valeurs

- Le problème du bon sens.

Donc même si vous précisez à la super IA “ne détruis pas toute vie sur Terre”, elle pourrait mettre en place une solution qui extermine entièrement les humains. Donc on obtient une source d'énergie propre et illimitée sans détruire toute vie sur Terre. Du point de vue de la super IA, elle a accompli son objectif. Pour nous, c’est moins drôle.

Si vous demandez à un comité d’expert de réfléchir à une source d'énergie propre et illimitée, vous partez du principe qu’ils ne vont pas trouver une solution qui va annihiler toute vie sur Terre ni exterminer l’humanité. Même si vous n’avez pas explicitement spécifié ces paramètres. Car entre humains, nous partageons du bon sens commun.

Il y a également le scénario où l’on demande à la super IA de résoudre un problème qui n’a pas de fin claire et précise. Par exemple construire des meilleurs processeurs pour augmenter la capacité de calcul de nos ordinateurs. Au début, la super IA va utiliser des nouvelles technologies et multiplier par 2 la vitesse de calcul 181

de nos ordinateurs. Super, tout le monde est content. Mais la super IA continue dans son objectif, car de son point de vue, il est encore possible de construire des meilleurs processeurs. Donc elle transforme les atomes composant les forêts en atomes utiles pour des super processeurs grâce à des technologies de nanoassemblage. Puis, les atomes des montagnes, plages, déserts, et tous les êtres organiques sur la planète. La Terre entière devient ainsi un super ordinateur. Mais pourquoi s'arrêter là ? Car du point de vue de la super IA, il est encore possible de construire des meilleurs processeurs. Elle convertit donc la Lune, Mars, les astéroïdes et tout le système solaire en super ordinateur. Objectif accompli ? Non, car il est encore possible de construire de meilleurs processeurs ...

Alors on pourrait se demander pourquoi cette super IA ne comprend pas quand s'arrêter. Elle agit finalement de façon débile pour une super intelligence ! Encore une fois, notre intuition sur ce qu’est l’intelligence nous trompe. Une super IA pense comme un ordinateur, car c’est un

ordinateur. Mais lorsque nous imaginons une IA hautement intelligente, nous commettons l'erreur de l'anthropomorphiser (projeter les valeurs humaines sur une entité non humaine) parce que nous pensons d'un point de vue humaine et que, dans notre quotidien actuel, les seules choses qui possède une intelligence de niveau humain sont des humains. Pour comprendre une super IA, nous devons comprendre qu'une entité peut être extrêmement intelligente et en même temps, totalement étrangère à notre intelligence.

Tim Urban, du blog Waitbutwhy.com, prend pour analogie l'intelligence d'une tarentule : "*Imaginez que vous avez rendu une araignée beaucoup, beaucoup plus intelligente, à tel point qu'elle surpasse de loin l'intelligence humaine ? Est-ce qu'elle deviendra tout à coup plus bénéfique pour nous et ressentira des émotions humaines telles que l'empathie, l'humour et l'amour ? Non, car il n'y a aucune raison que le fait de devenir plus intelligente la rende plus humaine. Elle sera incroyablement plus intelligente, mais restera aussi fondamentalement une araignée dans son fonctionnement interne. Et c'est extrêmement effrayant. Je ne voudrais pas passer du temps avec une araignée super intelligente. Et vous ?* "

C'est un peu la même idée avec une intelligence artificielle. En devenant plus intelligente, elle ne deviendra pas spécialement plus humaine que l'est votre ordinateur aujourd'hui.

Chez les humains, plus d'intelligence a été également accompagnée de motivation plus complexe. Nos motivations sont passés de se nourrir et se reproduire, à créer des liens sociaux, fonder des sociétés, bâtir des cathédrales, explorer les océans, découvrir les lois de la physique, etc. Mais Nick Bostrom met en avant le fait que l'intelligence et la motivation ne sont pas étroitement liées. Il appelle cela la thèse orthogonale.

C’est-à-dire qu’un système qui devient plus intelligent ne va pas forcément développer des objectifs plus complexes. Si son objectif initial lorsqu’il a une IA limitée, c’est de calculer les décimales de Pi. Il utilise un supercalculateur et reste inoffensif. Mais il pourrait garder cet objectif lors de son passage à une super intelligence. Ce qui rendra ce système extrêmement dangereux, car dans le but de calculer les décimales de Pi, il devra acquérir plus de puissance de calcul, et donc entreprendre de construire plus de supercalculateurs. Étant une super intelligence, il possède des compétences faramineuses et pourrait recouvrir la Terre entière de supercalculateur.

Nick Bostrom est à l'origine de la thèse orthogonale qui prétend qu'un système super intelligent peut poursuivre des objectifs super simples. Son exemple le plus connu est une IA qui fabrique des trombones, mais en devenant une super IA, elle garde cet objectif. Elle recouvre donc la Terre en usines à fabriquer des trombones.

©Future of Humanity Institute CC BY-SA 4.0

Une “super” intelligence possède des “supers” pouvoirs sur son environnement et par conséquent peut très rapidement entreprendre des changements sur nos modes de vie qui pourraient être désastreux si l’on n’a pas correctement spécifié ce que nous voulons.

Tout le monde ne partage pas l’idée qu’une super intelligence artificielle sera dangereuse. Alan Winfield, professeur à l'Université de Bristol, a déclaré que SI nous 183

parvenons à construire une IA équivalente à l'homme et SI cette IA acquiert une compréhension totale de son fonctionnement, et SI elle parvient ensuite à s'améliorer elle-même, et SI elle produit une intelligence artificielle super intelligente, et SI cette super-intelligence, par accident ou par malveillance, commence à consommer des ressources, et SI nous n'arrivons pas à l'arrêter, alors oui, nous pourrions bien avoir un problème. Le risque, bien qu’existant, est improbable, car il présuppose de nombreuses conditions qui sont loin d'être évidentes.

Steven Pinker, psychologue évolutionnaire pense qu’un scénario de type roi Midas ne tient pas debout. Selon lui, l’idée que l’on sera suffisamment intelligent pour réussir à créer une super intelligence artificielle, mais en même temps suffisamment stupide pour lui demander de résoudre un problème qui entraînera notre extinction est ridicule.

Même si de nombreux chercheurs pensent que de dresser des scénarios dystopiques sur une potentielle super intelligence artificielle n’est qu’une distraction, le philosophe Nick Bostrom a un autre avis. Selon lui, il vaut mieux passer du temps à réfléchir à tous les détails qui pourraient entraîner des conséquences désastreuses, plutôt que de dresser une liste de tout ce qui sera bénéfique. Car il vaut mieux être pris par surprise par une bonne application d’une super intelligence artificielle, plutôt qu’une mauvaise que l’on n’avait pas vue venir et qui met en péril l'espèce humaine.

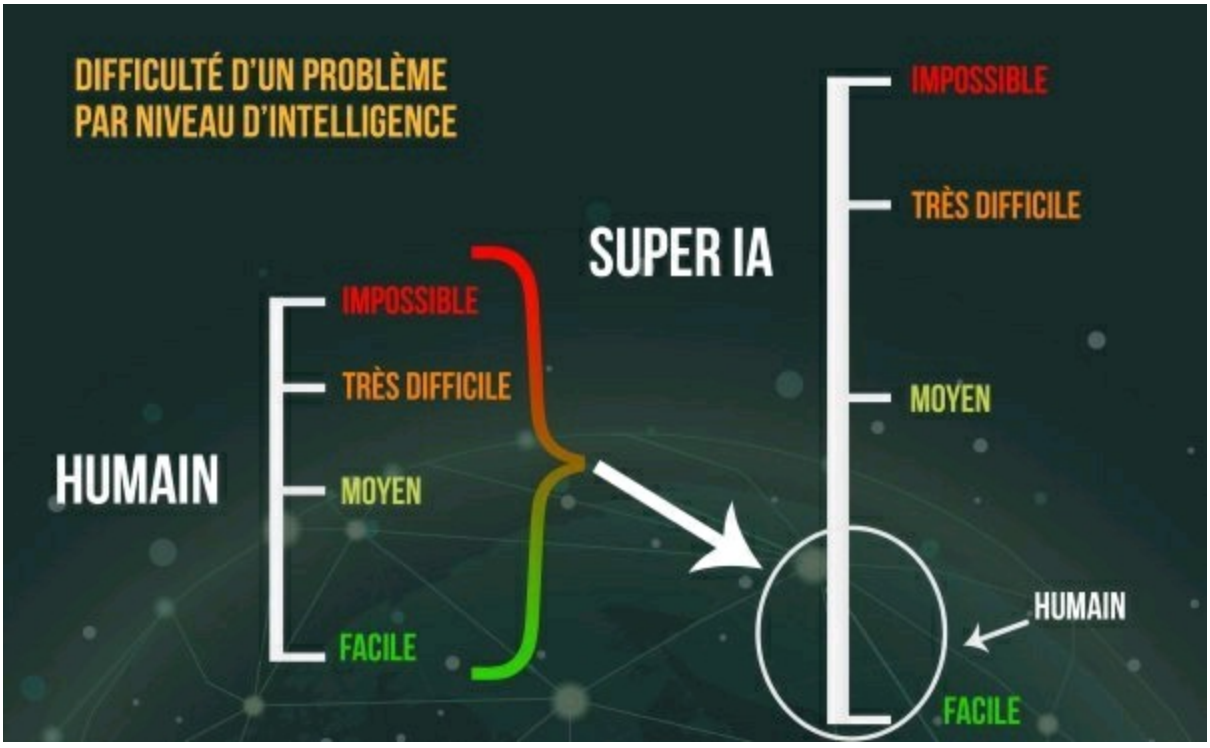
D’autant plus que ce n’est pas parce qu'on alloue beaucoup d'énergie à planifier les potentiels désastres, que ça les rend plus probables.

Si demain, nous captons un signal venant de l’espace disant : “*Nous sommes en route vers votre planète. Nous arriverons dans 30 ans.*

*Signé une civilisation extraterrestre.* ” Est-ce qu’il serait judicieux de se dire que cela ne sert à rien de s'inquiéter pour quelque chose qui arrivera dans 30 ans ?

L'être humain apprend souvent de ces erreurs. Sur le plan personnel, c’est lorsqu’on se brûle que l’on sait qu’il ne faut pas mettre la main sur la plaque de cuisson. C’est encore plus vrai avec nos technologies. On se rend compte que le feu peut détruire des maisons, donc on invente l’extincteur. C’est lorsqu’il y a un crash aérien que les compagnies changent leurs normes de sécurité. La ceinture de sécurité a été inventée après la mise en service des premières voitures. Le problème avec un risque existentiel, c’est que la première erreur est aussi la dernière et il n’y a plus personne pour apprendre quoi que ce soit.

184




4. Après demain: Super Intelligence Artificielle

La clé pour un futur viable

Le chercheur Eliezer Yudkowsky a déclaré : “*Aucun problème n’est*

*impossible à résoudre, il y a seulement des problèmes difficiles à résoudre pour un certain niveau d'intelligence. Déplacez le curseur un cran au-dessus, sur l'échelle d'intelligence, et certains problèmes vont soudainement passer de "impossible" à "évident". Déplacez le curseur de 10 crans au-dessus, et tous les problèmes deviennent évidents. "*

Un problème difficile pour un certain niveau d'intelligence est trivial pour un niveau d'intelligence plus élevé. ©Gaëtan selle

**Résoudre nos problèmes les plus difficiles**

L'intelligence est un moyen très puissant pour résoudre des problèmes. Nous avons beaucoup de problèmes à résoudre aujourd'hui, donc logiquement, nous avons besoin de plus d'intelligence. Une super intelligence artificielle a clairement le potentiel de nous aider à résoudre nos problèmes. Voilà en résumé l'idée que mettent en avant les technos optimistes.

Et si l'on compare cela avec notre propre capacité à résoudre les problèmes d'une 185

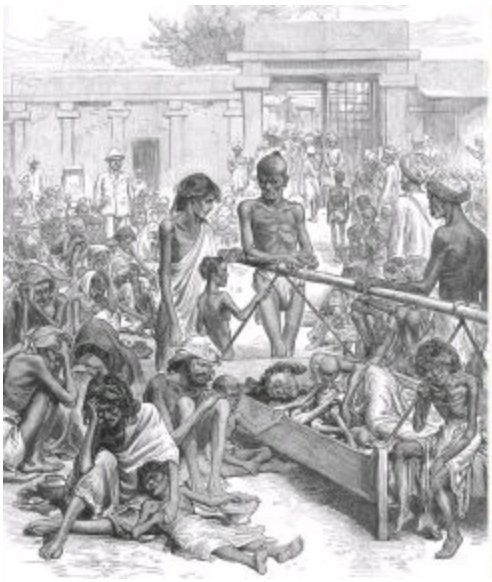
autre espèce, on comprend que notre relation avec une super intelligence artificielle peut être extrêmement bénéfique. Par exemple, un chien est un animal qui fait face à plusieurs problèmes : Trouver de la nourriture, rester en bonne santé, survivre, jouer avec d'autres chiens, se socialiser, se balader en toute sécurité. Les chiens sont clairement des animaux intelligents, mais en s'alliant avec une intelligence supérieure, celle des humains, ils n'ont plus à se soucier de ces problèmes. Ils sont nourris sans avoir à chasser, ils sont soigner et vacciner pour des maladies qu'ils ne savent même pas comment ni par quoi elles sont causées, ils sont promenés dans des parcs où ils peuvent s'amuser avec d'autres chiens. On peut se dire que la qualité de vie d'un chien appartenant à un humain est supérieure à celle d'un chien errant. Et généralement, leur espérance de vie aussi. Sauf cas extrême où le chien est battu par son maître.

C'est idée est notamment mise en avant par Elon Musk qui pense que l'un des meilleurs scénarios possible après l'émergence d'une super IA, c'est qu'elle nous considère comme des animaux de compagnie. C'est

selon lui, une meilleure alternative que l’extinction.

Il ne faut pas oublier que nous créons de la technologie dans le but de nous aider à régler nos problèmes. Et cela marche puisque la technologie nous a permis de mettre fin aux deux plus gros problèmes qu'a fait face notre espèce depuis le début de son histoire. Effectivement, si vous ouvrez des livres d'histoire, vous constaterez deux obstacles récurrents qui ont eu la fâcheuse tendance à se mettre sur notre chemin. La famine et les épidémies. Nos ancêtres ont prié un nombre considérable de Dieu, Déesse, anges et esprits dans le but d'obtenir de l'aide contre ses deux tourments. Ça n'a pas marché. Nous avons essayé toute sorte de systèmes politiques et modèle économique. Ça n'a pas marché non plus. Mais au cours des deux derniers siècles, en grande partie grâce à l'ingéniosité de notre espèce et au développement scientifique et technologique, nous avons pris le contrôle sur la famine et les épidémies.

La famine a accompagné l'humanité depuis le berceau de la civilisation. Jusque récemment, la majorité des humains vivaient dans des conditions très rudes où la moindre erreur ou un peu de malchance pouvait facilement être synonyme de peine de mort pour le village tout entier qui se retrouvait sans nourriture. Lorsque la sécheresse frappait l’Inde ou l’ancienne Égypte, il n'était pas rare que 5 à 10 % de la population périssent. Les provisions étaient trop justes, le transport trop lent et cher et les gouvernements étaient bien impuissants face aux caprices de la nature. Mais aujourd'hui, grâce au développement technologique, il y a plus de personnes qui meurent de trop manger que de ne pas assez manger. Et ça, c'est vraiment un fait qui était inimaginable il n'y a ne serait ce que un siècle. Si en 2019, un être humain meurt, car il n'a pas eu assez à manger, en Syrie, en Corée du Nord ou en Somalie, ce n'est 186



pas dû à une cause naturelle comme c'était le cas dans la passé. Que ce soit une invasion de sauterelles, un hiver rude ou une maladie dans les champs. Mais c'est bien parce qu'un leader politique, un gouvernement ou une idéologie a permis à cet être humain de mourir de faim. En vérité, trop manger est devenu un problème bien plus grave que la famine. C'est dire l'évolution !

Au niveau des épidémies, c'est pareil. Il fut un temps où les gens vivaient leur vie en se disant que la semaine prochaine, il pourrait attraper une maladie et mourir presque sur-le-champ. Ou qu'une épidémie pourrait éradiquer leur famille entière en quelques jours. Lors des derniers siècles, l'humanité s'est rendue encore plus vulnérable aux épidémies grâce à un large réseau de transport international et une démographie en plein boom. Pourtant, les épidémies ont baissé drastiquement. Ce paradoxe s'explique par des progrès ahurissants en médecine qui se traduisent par la vaccination, les antibiotiques, une meilleure hygiène et des infrastructures médicales de qualité. Nous ne sommes évidemment pas à l'abri d'une épidémie, mais aujourd'hui, la médecine a tendance à gagner la course contre la nouvelle infection.

On peut donc dire que les épidémies sont des risques bien plus faibles pour l'humanité qu'elles ne l'étaient par le passé. Nous mourrons bien plus de maladie non infectieuse comme le cancer. En faite, il y a largement plus de personnes qui meurent de vieillesse que de personne mourant de

maladie infectieuse. C'est aussi une première dans notre histoire !

La famine et les maladies infectieuses sont désormais bien plus sous contrôle grâce à la science et la technologie.

Tout cela pour dire que l'intelligence artificielle est également une technologie qui a pour but de nous aider à régler des problèmes. Et lorsque l'on aura une super IA, elle sera super compétente pour nous aider (si elle est sous contrôle).

187

Si l'on compare ce qu'est capable l'intelligence humaine, comparée à ce qu'est capable l'intelligence des autres animaux, on arrive plus facilement à saisir l'implication qu'une super IA pourrait avoir face à nos problèmes. En effet, il est difficile de trouver ne serait-ce qu'un seul problème difficile pour, disons un chimpanzé, qui serait également un problème difficile à résoudre pour nous. Si on imagine un groupe de chimpanzé demander notre aide pour se débarrasser d'un prédateur leur posant des soucis au quotidien, on n'aura pas trop de mal à trouver une solution. Si une colonie de fourmis souhaite régler leur problème de guerre constante avec une fourmilière voisine, ce sera facile pour les humains. On pourrait les enflammer, les noyer, tuer leur reine, perturber leur phéromone, déplacer complètement la fourmilière … les solutions sont nombreuses, triviales et complètement hors de porté de la plus intelligente des fourmis. Et si on le décide, on pourrait créer un véritable paradis pour une fourmilière, en leur donnant tout ce dont elles ont besoin pour s'épanouir pleinement. On ferait office de super intelligence pour les fourmis.

Par analogie, une super intelligence artificielle possédera tellement de capacité cognitive que nos problèmes les plus complexes seront triviaux pour elle. Lorsqu'un enfant de 5 ans vous montre ses devoirs d'école, vous n'avez généralement pas trop de problèmes à les résoudre. C'est trivial pour votre intelligence. Et si un problème comme générer de l'énergie propre et renouvelable à l'infini était également trivial pour une super intelligence ?

Le philosophe Nick Bostrom pense qu'il est difficile de penser à un

problème qu'une superintelligence ne pourrait pas résoudre, ou du moins, nous aider à résoudre.

Maladie, pauvreté, destruction de l'environnement, souffrances inutiles ... voilà ce qu'une superintelligence dotée de technologies de pointe serait capable d'éliminer.

La tendance naturelle de l'être humain est de rejeter l'idée que les problèmes majeurs d'aujourd'hui vont disparaître sur le court terme. Comme la pauvreté, les maladies, le réchauffement climatique, la surconsommation des ressources naturelles.

On considère ces problèmes comme faisant partie du paysage, et qu'ils vont rester pendant encore longtemps. Pensez que le quotidien est normal, est au final inscrit dans nos gènes. C'est une façon de s'adapter à l'environnement. Imaginer une créature qui serait traumatisée par le monde qui l'entoure en permanence. Pas idéal pour survivre et transmettre ses gènes. Et le faite qu'on évolue depuis la naissance dans un monde en constant changement nous incite à ne pas trop nous poser de question. Car nos ancêtres de l'âge de pierre ont vécu pendant des dizaines de milliers d'années dans des environnements similaires, ou chaque génération vivait plus ou moins la même vie. On est génétiquement programmé pour penser que le monde est normal. Que nos parents vivaient une vie plus ou moins semblable avec quelques



gadgets technologiques en moins, et que de nos enfants vivront une vie plus ou moins semblable, avec quelques gadgets technologiques en plus. Penser que la prochaine décennie sera complètement différente va à l'encontre de notre expérience de tous les jours, et à l'encontre de la programmation génétique.

Pourtant, les choses changent et évoluent. Si on pouvait montrer à un chasseur cueilleur de l'âge de pierre, notre époque d'abondance, de confort et de loisir, il n'y verrait que de la magie qui ne fait aucun sens. Le monde qu'une super IA pourrait créer a le potentiel d'être tout aussi magique et incroyable.

Comme le mathématicien Irvin J. Good (1916 - 2009) l'a si bien dit, " *une super intelligence artificielle sera la dernière invention de l'humanité. Car une fois en place, elle sera la source de nos innovations.* " Toutes les technologies en cours de recherche ou qui nous manquent pour résoudre nos problèmes les plus difficiles pourraient être drastiquement accélérées, voire même complètement résolues par une super IA.

Le réchauffement climatique est surement dans le top 3 de la liste. Une super IA aurait la capacité de comprendre le problème avec un degré de précision inimaginable, et appliquer des solutions appropriées pour tout d'abord, mettre en place des techniques de production d'énergie 100% renouvelable, propre et illimitée.

Fusion nucléaire, satellite solaire, une technique inconnue... Si l'univers est infini, il doit bien y avoir une quantité infinie d'énergie. Ensuite, développer des technologies avancées pour enlever les excès de $CO_2$ dans l'atmosphère. On a déjà des pistes pour faire cela comme avoir plus de forêts (peut-être dans les régions désertiques), ou via des procédés chimiques impliquant de l'oxyde de calcium ou hydroxyde de sodium.

Donc une super IA trouvera surement des moyens encore plus efficaces.

Quoi d'autre dans la liste … ah oui, la production et distribution alimentaire. Déjà, le fait que certaines personnes meurent encore de faim aujourd'hui est inadmissible.

Ensuite, la qualité de la nourriture que l'on mange est loin d'être idéale. Entre pesticides, antibiotiques et toute la pollution générée par la production. Une super IA pourrait concevoir un assembleur moléculaire afin de créer de la nourriture à partir d'élément de base, et la distribuer à toute la planète via des moyens de transport qui dépasse de loin nos engins les plus rapides.

Cette même technologie d'assembleur moléculaire est également la clé pour concevoir n'importe qu'elle objet à partir de ses atomes de base. Il n'y aura donc plus besoin de les construire "à l'ancienne". L'idée de main-d'oeuvre sera obsolète et il est difficile d'envisager à quoi pourrait ressembler l'économie dans un tel futur.

Une super IA pourrait trouver de meilleures solutions politiques, sociales et économiques afin de mieux coopérer sur la planète. La démocratie fonctionne mieux si tout le monde vote en faveur de ce qui est le mieux pour le pays dans son ensemble.

Malheureusement, en réalité, nous avons tendance à voter en faveur de ce qui est le mieux pour nous-mêmes. Une majorité de la population peut totalement gouverner une minorité de la population, simplement parce qu'elle est majoritaire lors des élections. Avec une super IA, nous aurons une bien meilleure option. Elle peut simplement exécuter chaque décision politique sur une simulation complète de l'ensemble de la population grâce à son immense capacité de calcul, et déterminer les meilleures décisions politiques et législatives. Tout comme dans le domaine de l'architecture, ou de l'aérodynamisme, tester des paramètres spécifiques sur des simulations est le meilleur moyen d'assurer de meilleurs résultats sans les conséquences d'un échec dans le monde réel.

Et bien sûr, une super IA nous aidera à éliminer les maladies qui affligent de terrible souffrance comme le cancer, alzheimer, parkinson, les maladies infectieuses, etc. Ce dernier point nous amène à sérieusement envisager qu'une super IA détiendra peut-être la réponse à une question que de nombreux chercheurs sur la planète tentent de trouver :

**Peut-on mettre fin à notre mortalité biologique ?**

Le vieillissement est de plus en plus considéré comme un processus qui s'apparente à une maladie, et donc qui a la possibilité d'être soigné. Les chercheurs travaillent sur plusieurs pistes et ont déjà réussi à allonger la durée de vie de plusieurs espèces, dont les rats. Un traitement contre le vieillissement humain n'est qu'une question de temps. Et une super IA pourrait mettre sa super intelligence au service de cette cause.

Déjà, comme on l'a vu dans la deuxième partie, nos IA personnelles vont nous aider à rester en bonne santé. Des capteurs dans votre maison analyseront en permanence votre souffle pour détecter les premiers signes de cancer et des nanorobots nageront dans votre

circulation sanguine pour dissoudre les caillots sanguins dans votre cerveau ou vos artères avant de causer un accident vasculaire cérébral ou une crise cardiaque. Votre IA personnelle servira d'assistant médical 24h /

24 et 7j / 7. EIle surveillera vos réponses immunitaires, vos nutriments et votre métabolisme, en développant une vision à long terme de votre santé qui donnera aux médecins une idée précise de ce qui se passe dans votre corps.

Une super IA pourrait également nous aider a révolutionné la modification de notre génome afin de modifier l'ADN humain de la même manière qu'un éditeur 190

corrige un mauvais manuscrit, coupant les sections problématiques et les remplaçant par des gènes bénéfiques. Seul un système super intelligent pourrait cartographier l'interaction extrêmement complexe des mutations géniques et les mettre en relation avec ce que l'on désire préserver ou modifier. Elle saura donc ce qui nous permet de vivre plus longtemps, et ce qui dégrade notre santé.

Avec sa super intelligence et capacité de calcul, une super IA pourrait également scanner complètement notre cerveau pour le numériser. Nous ne savons pas si une telle chose est possible ou pas, mais l'hypothèse n'est pas à écarter. Dès lors, nous pourrions évoluer dans des mondes simulés indissociables de la réalité. Les types d'expériences que nous pourrions avoir n'ont pour seule limite que notre imagination.

Une super IA pourrait également nous donner la possibilité d'augmenter considérablement nos capacités intellectuelles et affectives et nous aider à créer un monde dans lequel nous pourrions vivre des vies consacrées à tout ce qui nous apporte de la joie. En d'autres termes, un paradis sur Terre. D'autres voient cela plus comme un zoo pour les humains où tous nos besoins sont pris en charge. Mais au final, peu importe le terme que l'on utilise, le monde d'abondance dans lequel nous pourrions évoluer après l'émergence d'une super intelligence artificielle paraît tout simplement irréaliste. Tout autant que l'est notre époque pour quelqu'un vivant au moyen âge.

Sans oublier que les nouvelles technologies développées par une super IA nous aideront à nous répandre dans l'espace. Propulsion révolutionnaire, terraforming, cryonie, colonisation et exploration de la galaxie, système solaire après système solaire. Le terrain de jeu pour une espèce intelligente est astronomiquement gigantesque. Il est estimé que 107 milliards d'humains ont vécu jusqu'à ce jour. Mais ce nombre est ridicule comparé au futur possible de l'humanité. Si l'on devient une civilisation galactique, on parle de quintillions d'humains, minimum ...

Vivre plus longtemps, être immortel, guérir toutes les maladies, développer des énergies propres et illimitées, créer une abondance pour tous, coloniser les planètes du système solaire, créer des réalités virtuelles indissociables de la réalité ... Au final, tout ce qui n'est pas en contradiction avec les lois de la physique est envisageable sous la condition de posséder la connaissance adéquate.

Mais quelle forme pourrait prendre cette super intelligence ? Voici trois types de super intelligence artificielle proposés par le philosophe Nick Bostrom sans son livre

"Superintelligence".

**Un oracle**

191

Il s'agit d'un système super intelligent, conçu pour répondre uniquement aux questions et ne pouvant pas agir dans le monde. Une oracle super IA pourrait par exemple recevoir des questions sous formes écrites ou orales, et nous donner des réponses de manière analogue. D'une certaine façon, une calculatrice est une forme primitive d'oracle, et Google également.

L'oracle super IA doit être capable de comprendre le langage et les concepts humains de façon précise et détaillée. Elle devra également d'une manière ou d'une autre avoir accès à de l'information sur le monde réel et à la connaissance humaine si on souhaite qu'elle nous aide à résoudre nos problèmes. Ce qui implique d'être connecté à Internet par

exemple, ce qui pose son lot de danger, car nous ne voulons pas que l'oracle super IA puisse modifier du contenu sur Internet. Il faudrait qu'elle soit en "lecture seule" par exemple.

L'exigence la plus importante pour une oracle super IA est qu'elle soit précise, autant que ces capacités le permettent au moment où on lui pose une question. Car nous ne voudrions pas, par exemple, que l'oracle super IA utilise ses super compétences pour prendre le contrôle des ressources du monde réel afin de construire une meilleure version d'elle-même qui répondra mieux à la question. Elle devra faire avec les ressources dont elle dispose pour nous donner la réponse la plus précise possible.

On peut poser deux types de questions à une oracle super IA : Des prédictions et des problèmes. En général, les prédictions sont des questions du type "*que se passerait-il si…?*", alors que les problèmes sont du type "*comment pouvons-nous résoudre ceci ou cela …?*", bien qu'il existe un certain chevauchement entre les catégories (on peut résoudre des problèmes grâce à une utilisation intelligente de plusieurs prédictions, par exemple). Mais dans les deux cas, il existe de grands risques sociaux si l'on pose certains types de questions. Nous vivons dans un monde polarisé entre des différences politiques, nationalistes et religieuses. Entre des personnes occupant des positions de pouvoir et de richesse considérable qui veulent s'accrocher à ces privilèges, et des gens qui les veulent. Certaines questions pourraient renverser ces hiérarchies, et devenir source de conflit majeure. Ainsi, à moins que l'oracle super IA soit entièrement secrète, les concepteurs doivent s'engager à ne poser aucune question qui leur donneraient un grand pouvoir au détriment des autres. Ni des questions qui iraient au cœur de puissants mouvements idéologiques (tels que "*Y a-t-il un dieu ?*" ou "*quelle culture est la meilleure ?*"). Un monde post- super IA va être très différent du nôtre, nous devons donc agir pour minimiser les perturbations initiales, même si cela implique de laisser certaines questions sans réponses.

Différents moyens peuvent être envisagés pour éviter ces questions perturbatrices.

Les concepteurs de l'oracle pourraient publier, à l'avance, une liste de toutes les questions que l'on pourrait poser. Un vote démocratique sur Internet serait pris, avec des questions nécessitant une supermajorité (par exemple, approbation à 90%) avant d'être soumises à l'oracle super IA. Ou on pourrait demander à l'oracle elle-même de trancher, en terminant chaque question par une mise en garde du type "*ne réponds pas à cette question si la réponse est susceptible d'avoir un impact négatif sur le monde*".

Pour prendre des mesures de précautions contre un résultat désastreux, il est préférable de poser des questions dont nous pouvons vérifier les réponses. Également des problèmes dont les solutions sont réversibles - en d'autres termes, si nous appliquons la solution qu'une oracle super IA nous a conseillé, et que ça tourne mal, nous devrions pouvoir inverser ses effets. Par exemple : "*Quels matériaux pourrait-on utiliser pour construire des avions plus rapides et légers ?* " serait un problème idéal à poser. La solution de l'oracle super IA est susceptible d'être tout à fait compréhensibles, réversibles et peut-être même vérifiables par des ingénieurs. Par contre si on lui demande : "*Trouve une nouvelle façon de gérer l'économie mondiale*

*?* " et en appliquant aveuglément sa suggestion à grande échelle, cela pourrait conduire à une catastrophe.

Des problèmes spécifiques ont également l'avantage de réduire les dangers sociaux des questions posées. Peu de gens dans le monde s'opposeraient fermement à des questions du type "*Comment pouvons-nous soigner le SIDA / le cancer / la tuberculose ?* ", ou "*comment pouvons-nous concevoir de meilleures batteries électriques ?* " ou " *quelles seront les conséquences d'une augmentation de 1% de l'impôt sur le revenu dans l'économie pour les dix prochaines années ?* "

**Un génie**

Un génie super IA reçoit une commande précise, et à la possibilité de l'exécuter jusqu'à ce que l'instruction soit achevée. C'est un oracle qui n'est plus simplement cantonné à répondre à des questions, mais qui peut interagir dans le monde pour les résoudre. Cela implique donc que le système super intelligent a la capacité de donner des instructions à des machines

reliées à son réseau. Par exemple des machines industrielles, des imprimantes 3D, etc. Une usine entière pourrait être sous son contrôle le temps que le génie super IA accomplisse la commande donnée par les humains. En reprenant l'exemple : "*Quels matériaux pourrait-on utiliser pour construire des avions plus rapides et légers ?* ", on lui demandera plutôt "*Construit un avion plus rapide et léger*". Le génie super IA aurait non seulement la solution sur papier, mais exécutera la tâche en produisant un avion plus rapide et léger dans une usine qu'il contrôle. Il peut aussi donner des ordres à des ingénieurs et ouvriers humains pour fabriquer certaines parties spécifiques.

193

Si l'on créait un génie super IA, il serait souhaitable de le construire de façon à ce qu'il obéisse à l'intention de la commande plutôt qu'à sa signification littérale ou il pourrait devenir un risque existentiel pour des raisons que l'on a vues dans le chapitre approprié. Dans le but de limiter les conséquences désastreuses, une option serait de créer un génie super IA qui présenterait automatiquement à l'utilisateur une prédiction des résultats probables d'une commande proposée sous forme de simulation, demandant confirmation avant de procéder. Une sorte de prévisualisation. Ainsi, si on lui demande de construire un avion plus rapide et léger, il pourrait produire une vidéo 3D photoréaliste de toutes les étapes qu'il s'apprête à faire dans la construction de l'avion. Si nous sommes satisfaits, on lui donne le feu vert. Si on juge qu'il y a des dangers, on annule la commande.

**Un souverain**

Selon la définition du Larousse : "*Se dit d'un pouvoir qui n'est limité par aucun autre*" ou encore "*Qui émane d'un organe souverain et n'est susceptible d'aucun contrôle*". À l'image d'un génie, un souverain super IA peut interagir dans le monde réel, mais la différence résulte dans l'ampleur de son intervention. Un génie super IA reçoit uniquement des commandes spécifiques, alors que le souverain reçoit des tâches qui n'ont pas nécessairement de solutions uniques, objectives, claires et précises.

Il a la possibilité de fabriquer de nouveaux objets, inventer de nouvelle

technologie voire même instaurer de nouvelles lois afin de poursuivre son but. Si on reprend l'exemple précédent, on lui demanderait : "*Invente un nouveau moyen de faire voler des humains rapidement sur de longues distances à travers la planète et avec un niveau de sécurité supérieure ou égale à nos avions actuels*". Avec cet ordre, le souverain a la liberté de s'y prendre comme il veut jusqu'à ce que l'on obtienne ce que l'on a demandé.

Si une super IA souveraine est bienveillante, cela pourrait être très positif pour le futur de la gouvernance humaine. On peut imaginer une telle IA souveraine qui se fond dans le décor et qui tenterait d'arranger les choses pour que la satisfaction des humains soit maximisée. On aurait finalement une sorte de Dieu quasi omniscient et omnipotent chargé de faire ce que l'on désire dans le meilleur de nos intérêts.

Pour résumer très rapidement la différence entre les types de super IA, imaginons que vous êtes une super IA pour un enfant de 5 ans. Ce dernier a faim et souhaite manger un repas. Si vous êtes un oracle, vous lui donner la recette des pâtes au gratin.

Si vous êtes un génie, vous cuisinez les pâtes au gratin. Et si vous êtes un souverain, vous inventez une machine qui s'occupera de lui cuisiner des repas sur mesure pour le reste de sa vie s'il le souhaite.



La vraie différence entre les trois types de super intelligence artificielle ne réside donc pas dans les capacités ultimes qu'elles acquièrent. Au lieu de cela, la différence réside dans le problème de contrôle. Une oracle super IA est confiné et n'a qu'un accès limité au reste du monde. Il semble donc raisonnable de conclure qu'il sera plus simple de contrôler un oracle plutôt qu'un génie ou un souverain. Un oracle permettra également une transition plus progressive vers l'ère des super intelligences artificielles.

L'idée d'être accompagné par une super intelligence bénéfique qui souhaite le meilleur pour nous n'est pas si étrangère que cela. En effet, la majorité d'entres nous avons vécu un moment dans notre vie où une intelligence supérieure veillait sur nous.

On l'appelait “papa et maman”. Et dans la plupart des cas, les valeurs morales comme interagir avec les autres gentiment, ne pas blesser autrui, respecter les autres, être généreux, etc. sont passées d’une génération à l’autre. Ce sont des valeurs que nous avons choisies collectivement, car elles aident une société à s'épanouir. Et c’est presque un miracle que nous arrivions si bien à transmettre les valeurs à tous les enfants de la planète. En y regardant de plus près, chaque parent n'écrit pas la liste des valeurs qu’il souhaite donner à son enfant à ne partir de rien. Ils se contentent de transmettre ce qu’ils ont eux-mêmes appris de leurs parents. Et les enfants sont généralement capables d’accepter ces règles et de les appliquer lors de leur croissance vers l'âge adulte. C’est peut-être donc un modèle à suivre pour enseigner à une super IA à devenir bienveillante.



4. Après demain: Super Intelligence Artificielle

Post humanisme et futur lointain

Depuis le début de ce livre, nous étions plutôt concernés par le futur proche, et à moyen terme. Mais ici nous allons faire un bond dans un futur plus lointain. Car plus les années passent au court de ce 21e siècle, plus il apparaît évident que ce siècle est un tournant majeur pour Homo Sapiens. Non seulement le poids des risques existentiels se multiplie lié au développement technologique. L’intelligence artificielle n’étant que l’un d'entre eux. Ce qui pose forcément une très grande responsabilité vis-

à-vis des futures générations. Si l’on ne passe pas le cap du 21e siècle, c’est tout simplement la fin du futur de l’humanité. Mais également la nature même de notre espèce va peut-être changer. De plus, le futur cosmique de l’humanité est surement en train de se jouer durant ce siècle.

Décidément, c’est excitant de vivre à cette époque !

**Post-humanisme**

Pendant quatre milliards d’années, toutes les formes de vie étaient sujettes

aux lois de la sélection naturelle et de la biochimie organique. Peu importe si l'on parle d'une tomate, d'un T rex, d'un plancton ou d'un homo sapiens. Mais depuis quelques décennies, la Science est en train de créer une nouvelle ère. Une ère où les formes de vie n'évoluent plus par la sélection naturelle, mais par le biais d'une intelligence. Une ère où les formes de vie s'émancipent des limites de la matière organique. Pour la première fois en près de quatre milliards d'années, nous allons voir des formes de vies inorganiques évoluer sur la planète. On vit vraiment un tournant dans l'histoire de notre espèce, de la vie et même de l'univers si la Terre est la seule planète habitée ! Et bien sûr, cela aura des implications énormes sur notre civilisation, notre économie, notre culture et au final, sur tout ce qui fait ce que l'on est aujourd'hui.

Nous sommes probablement une des dernières générations d'Homo sapiens. D'ici un siècle ou deux, la Terre sera dominée par des entités qui seront aussi différentes de nous, que nous le sommes des chimpanzés. Les principales marchandises de l'économie de la fin du 21e siècle ne seront pas le textile, les voitures ou les armes, mais les organes, les corps, et les cerveaux.

Alors tout cela est bien beau, mais en quoi est-ce que ça a un rapport avec l'intelligence artificielle, qui est le sujet de ce livre. Comme on l'a vu dans les 196

chapitres précédents, une super intelligence artificielle sera peut-être la dernière invention que nous aurons à faire. Si elle est bénéfique, alors elle développera les nouvelles inventions selon nos désirs. En regardant les différentes disciplines technologiques d'aujourd'hui, on réalise qu'une super IA pourrait stimuler drastiquement chacune d'entre elles. Les disciplines comme la robotique, le génie génétique, la cybernétique, les interfaces cerveau/machine, la réalité virtuelle ou encore la nanotechnologie pourraient voir leurs statuts des recherches être précipités en quelques années et faire des percées majeures. C'est ce rythme du changement, cette accélération des avancées technologique qui pourrait mener à l'émergence de post-humain et qui fait figure de scénario possible après la singularité technologique.

Un élément commun de la science-fiction est que les humains vont fusionner avec les machines, soit en améliorant technologiquement les corps biologiques grâce à la cybernétique. Ou en téléchargeant directement nos esprits dans des machines ou dans des environnements virtuels. Selon les penseurs transhumanistes, un post-humain est un être hypothétique dont les capacités fondamentales dépassent tellement celles des humains d'aujourd'hui qu'elles ne sont plus humaines selon nos normes actuelles. Le

transhumanisme se concentre sur la modification de l'espèce humaine via tout type de technologie émergente, y compris le génie génétique, le numérique, l'intelligence artificielle et la cybernétique. Un post-humain est généralement un être capable de se connecter au cloud par la pensée ou de transférer sa conscience dans des simulations.

De posséder des prothèses avancées ou carrément un corps entièrement robotique. De ne plus être sujet au vieillissement biologique, et donc à la mortalité. De posséder des nanorobots dans son corps le réparant continuellement. Entre autres.

197

Le futurologue et entrepreneur Ray Kurzweil est un fervent partisan de la fusion entre l'homme et la machine. ©Was a bee CC BY-SA 2.0

Par contre, "Post humain" ne fait pas nécessairement référence à un futur où les humains sont éteints. Comme avec d'autres espèces qui se différencient les unes des autres, les humains et les post-humains pourraient coexister sur la planète. Il existe bien sur des scénarios dystopiques sur ce sujet. Par exemple le cas où seule une élite des plus fortunées de la planète arrive à augmenter leur capacité en devenant post-humaine et finit par dominer les humains "traditionnels". Je trouve personnellement ce scénario peu probable dans la mesure où si c'est une super IA qui est à l'origine des percées technologiques, elles seront probablement peu chères et accessibles à tous, car elle aura trouvé des moyens efficace et rentable de produire les technologies.

Et la tendance actuelle qui résulte du développement technologique c'est ce que l'on appelle la démonétisation. La nourriture coûte moins cher, car il existe des technologies pour produire de façon plus rentable. L'énergie coûte moins cher, car de nouvelles technologies l'a rendu plus rentable à produire. Cette tendance s'applique à de nombreux domaines, comme l'a montré l'entrepreneur Peter Diamandis.

Un scénario qui me semble plus probable c'est que la transition entre humain et post-humain se fera par l'adoption progressive des nouvelles technologies. Tout comme les smartphones ou internet se sont

progressivement répandus sur le globe.

Les avantages de devenir post-humain seront trop grands pour être reniés. Certaines personnes n'accepteront pas de se faire implanter une puce dans le cerveau pour se connecter à internet par la pensé, d'autre ne voudront pas de membres artificielles ou être modifié génétiquement. Mais à terme, ils seront une minorité. Probablement appartenant aux plus anciennes générations, voir même vivant peut être dans des communautés isolées à l'image des amishs.

Une variante sur le thème post-humain est la notion de "dieu post-humain", popularisé par Yuval Noah Harari dans son livre "Homo Deus". L'idée que nos descendants, n'étant plus confinés aux paramètres de la nature humaine, pourraient devenir physiquement et mentalement si puissants qu'ils pourraient ressembler aux dieux selon les définitions mythologiques. Pendant des milliers d'années, les humains ont imaginé les dieux d'une manière bien spécifique avec tout un tas de particularités.

Et nous sommes en train d'acquérir petit à petit tous les aspects qui définissent ce que nos ancêtres considéraient comme appartenant aux dieux.

Jusqu'à présent, augmenter l'espèce humaine s'était amélioré nos outils extérieurs.

Mais prochainement, nous allons fusionner avec nos outils ce qui nous permettra de nous améliorer de l'intérieur. Le passage vers la divinité se fera en suivant 3 chemins.

Biologique, cybernétique et non organique.



La modification biologique vient de la découverte que le corps humain est loin d'avoir atteint son potentiel maximal. Pendant 4 milliards d'années, la sélection naturelle a bidouillé les corps biologiques. Des amibes aux reptiles, aux mammifères à Homo Sapiens. Et il n'y a aucune raison de penser que notre espèce est la dernière étape et que la vie sur Terre a atteint son apogée. Ce serait présomptueux. Quelques petites

modifications dans les gènes, hormones et neurones d'Homo Erectus ont suffi pour passer d'un animal produisant des cailloux tranchants à Homo Sapiens capable de produire des vaisseaux spatiaux et des ordinateurs. Qui sait ce qui pourrait être le résultat de quelques modifications supplémentaire. Mais cette fois-ci, on ne va pas attendre patiemment sur la sélection naturelle. Nous allons modifier le code du vivant ce qui pourrait créer une espèce aussi différente de nous que nous le sommes pour Homo Erectus.

L'étape cybernétique ira un cran au-dessus en fusionnant des parties synthétiques a un corps biologique. Implants, prothèses, millions de nanorobots, etc. ce qui permettra des capacités bien plus grandes que de simples modifications biologiques.

Et enfin la dernière étape sera de passer complètement la frontière du biologique pour s'aventurer dans les contrées inexplorées du synthétique. Grâce à l'intelligence artificielle, le transfert de conscience ou les réalités virtuelles. Après 4 milliards d'années dans le royaume du monde organique, la vie va se développer sur des substrats non organiques et prendre des chemins que l'on ne peut même pas concevoir dans nos rêves les plus fous.

Avec ces nouvelles capacités, nous aurons atteint la divinité. Alors cela peut paraitre excentrique et pas du tout scientifique. Car la pour la plupart des gens, la divinité est un concept métaphysique qui prend souvent la forme d'une figure omnisciente, paternelle et protectrice. Mais lorsqu'on parle de transformer les humains en des dieux, il faut plutôt comparer cela avec les dieux gréco-romains, égyptiens, vikings ou encore les devas indus. Nos descendants auront des défauts et des limitations, tout comme Zeus, Odin ou Osiris. Les dieux des mythologies mentionnés ne jouissaient pas d'omnipotence, mais plutôt de super pouvoir bien spécifique. Comme la capacité de détruire et créer la vie, de transformer leur propre corps, de contrôler leur environnement et la météo, de lire les pensés, communiquer à distance, voyager à des vitesses hallucinantes et bien sûr, de ne pas être mortel.

L'humanité est en train d'acquérir tous ses pouvoirs et en réalité, nous en avons déjà certains. Lorsqu'on prend l'avion ou que l'on téléphone à l'autre bout de la planète, nous ne pensons pas utiliser des pouvoirs divins.

Pourtant, c'étaient ce que faisant certains dieux des anciennes mythologies et nos ancêtres verraient cela comme étant divins.

199

Les post-humains pourraient avoir dépassé complètement les limites de la biologie grâce au génie génétique, à la cybernétique et à l'intelligence artificielle.

**Le futur cosmique de l’humanité**

Si le développement technologique actuel déclenche au cours de ce siècle une explosion d’intelligence, ce sera peut-être le début d’une ère cosmique pour l’humanité. Lorsqu’on imagine le futur lointain de notre espèce, on pense inévitablement à l’espace. En tous cas, c’est ce qui revient le plus souvent lorsque l’on regarde les oeuvres de Science Fiction. Et

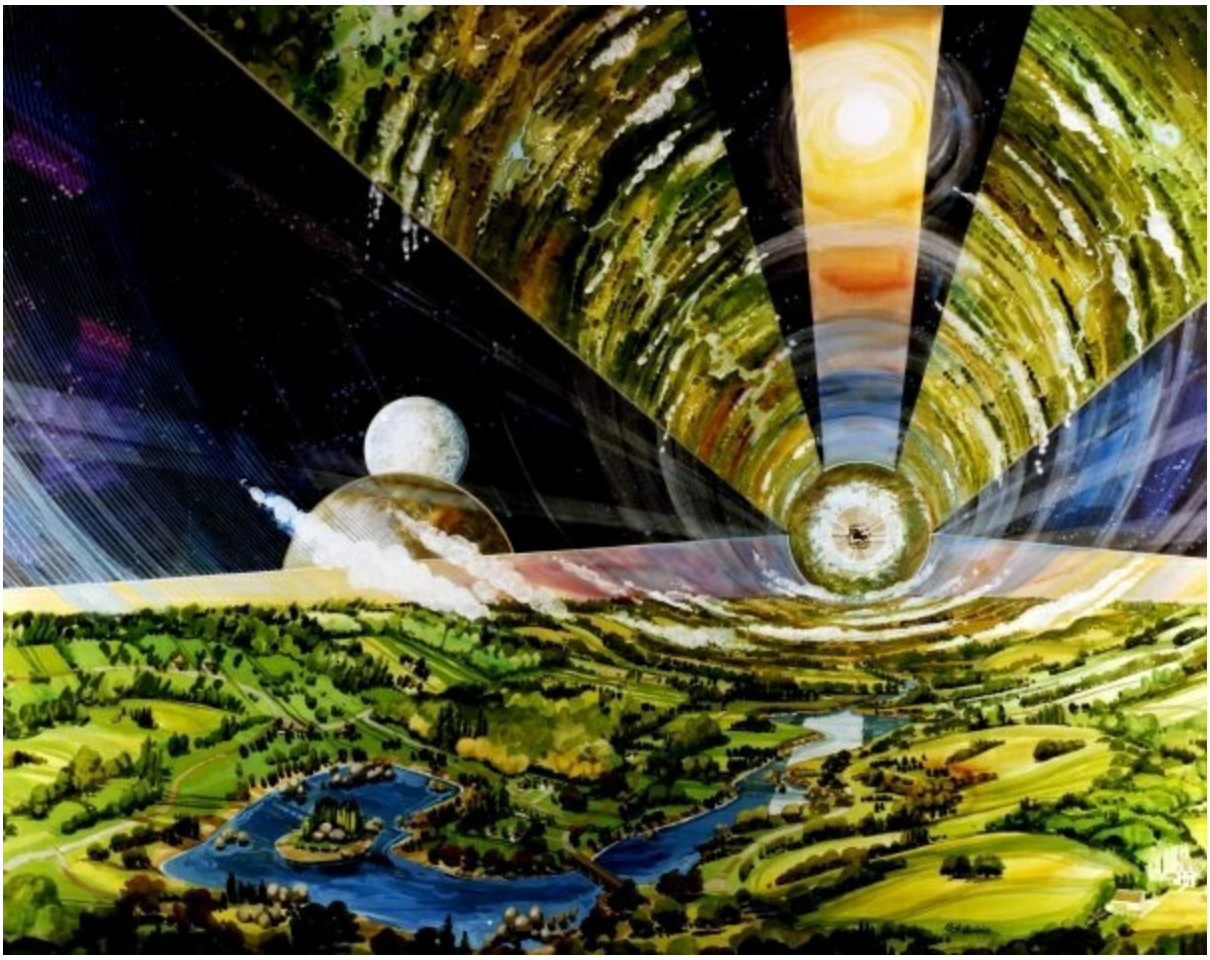
d’une certaine façon, c’est logique.

L’humanité est souvent décrite comme une espèce curieuse, motivée par l’exploration et l’envie de résoudre des mystères. L’espace est l’ultime frontière et si on ne colonise pas la galaxie, c’est forcément que l’on va s’autodétruire avant. Donc au final, prédire l’avenir est plutôt simple. Il n’y a seulement que 2 possibilités :

- L’humanité devient une civilisation galactique

- L’humanité disparaît.

De la saga Fondation d’Isaac Asimov, en passant par Dune de Frank Herbert, la série Star Trek, le jeu vidéo Mass Effect, Starcraft ou également l’idée d’empire galactique dans la saga Star Wars de George Lucas, la colonisation galactique est le 200



futur le plus populaire dans la science-fiction.

Si l'on part sur le principe qu'une super intelligence artificielle ne causera pas notre extinction, alors elle sera clairement un atout indispensable pour nous aider à répandre la civilisation dans la galaxie. Les lois de la physique, telle que nous les connaissons, imposent certaines limites. Une super IA nous permettra d'améliorer notre compréhension de l'univers et de ses lois, résultant ainsi peut être à dépasser les limites. Mais même en se basant sur les contraintes actuelles, comme la vitesse de la lumière, il existe des moyens théoriques pour se répandre à travers de vaste distance.

Les trous de vers étant le plus connus. Mais il y a également le moteur à distorsion (warp drive) qui distord l'espace, à la fois à l'avant et à l'arrière du vaisseau, lui permettant de voyager plus rapidement que la vitesse de la lumière. Et surement des moyens que nous ne connaissons pas encore, mais qu'une super IA pourrait découvrir et fabriquer.

Un cylindre d'O'neill est une des mégastructures qu'une civilisation post-humaine pourrait construire dans l'espace.

L'univers semble tristement vide de vie. Ce mystère est connu sous le nom du paradoxe de Fermi, qui postule que si l'univers est si grand, et ancien, nous devrions 201

voir des activités extra-terrestres de partout. Il devrait y avoir des milliers de civilisations dans notre galaxie. Se faisant la guerre, construisant des mégastructures, et visitant la Terre de temps en temps. Pourtant, nous ne voyons rien de tout cela ...

Nous devons donc prendre en compte le fait que nous sommes peut-être la seule espèce technologique et civilisationnelle dans la galaxie, voir dans tout l'univers. Ce n'est pas une théorie majoritairement acceptée par la communauté scientifique. La vérité c'est que nous n'en savons rien. Mais jusqu'à preuve du contraire, pour l'instant nous n'avons aucune preuve empirique que la vie s'est développée ailleurs dans le cosmos. Encore moins qu'il existe une civilisation extra-terrestre. Les observations d'OVNI sont au mieux des mystères, au pire des canulars. Il n'y

a que des hypothèses qui peuvent être proposées pour répondre au paradoxe de Fermi, et l'une d'entre elles c'est que nous sommes les seuls.

Cette hypothèse implique une énorme responsabilité morale ! Cela signifie qu'après 13,8 milliards d'années, la vie dans notre univers est à la croisée des chemins. Face à un choix entre s'épanouissant dans le cosmos ou s'éteignant sur une petite planète bleue. Laissant l'univers froid, inhabité et sans aucune expérience pour personne. Quel gâchis ! Si nous ne progressons pas technologiquement, la question n'est pas de savoir si l'humanité va disparaître, mais plutôt quand et comment. Qu'est-ce qui va annihiler la Terre en premier ? Un astéroïde, un super volcan, un rayon cosmique, la mort du soleil ? Autrement dit, sans la technologie, la fin de la vie sur Terre est imminente sur l'échelle cosmique. Il n'y a que le progrès technologique qui peut nous permettre de coloniser d'autres planètes et répandre la civilisation et la vie dans l'univers. Nous sommes peut-être les abeilles cosmiques, propageant les graines de la vie d'une planète à l'autre.

J'aime l'idée métaphysique que la Vie est le seul moyen qu'à l'univers d'être conscient de lui même. Et que la conscience en est le summum. Nous avons donc cette immense responsabilité de rendre l'univers encore plus conscient de lui même, en nous propageant le plus loin possible, pendant le plus longtemps possible.

Le célèbre astrophysicien Carl Sagan a déclaré que nous sommes à un stade d'adolescence technologique. C'est-à-dire un moment dans notre histoire où notre technologie nous rend plus puissants, mais notre sagesse n'est pas à la hauteur de cette puissance. C'est une position dangereuse qui peut conduire à notre extinction.

Tout comme il n'est pas sage de donner à des adolescents de 14 ans, une caisse de grenade. Mais la phase qui suit s'appelle la maturité technologique.

Il existe surement un plateau technologique. C'est-à-dire un seuil où les lois de la physique ne permettent plus à une branche spécifique de l'arbre technologique de 202

progresser. Par exemple, il existe peut-être une limite sur la quantité

d'énergie que l'on peut produire, et générer. Mais il est clair que, quel que soit ce plateau, il est très très loin au-dessus de ce que nous sommes capable aujourd'hui. Ce qui veut dire que nous avons un immense potentiel pour grandir en tant que civilisation. Et ce n'est pas la place ni les ressources qui manquent.

**Intelligence sans conscience**

Voici le scénario qui m'effraie le plus. Celui qui me semble le plus horrible selon tous les systèmes de mesure possible. Une super intelligence est développée sur Terre, elle échappe à notre contrôle, cause notre extinction et se répand dans la galaxie. Je ne suis pas nécessairement effrayé par notre extinction (même si ce n'est pas ma définition d'un futur bénéfique). Ce qui me semble tragique, c'est la perspective que cette super IA puisse ne pas être consciente.

En effet, nous voulons que nos descendants (IA, post-humains ou autre) soient conscients lorsqu'ils iront coloniser le cosmos sur les milliards d'années à venir. Sans conscience, il ne peut y avoir ni bonheur, bonté, beauté, valeurs, sens de la vie.

Comme je l'ai dit précédemment, la conscience est le moyen qu'à l'univers de s'expérimenter lui même. Sans conscience, c'est juste un gaspillage astronomique d'espace.

L'idée qu'une super intelligence artificielle aux compétences divines se répande dans l'univers avec des objectifs inconnus est terrifiante. Car elle aura besoin de collecter des ressources pour continuer son objectif. Et donc de détruire des planètes et des étoiles. S'il existe d'autres formes de vie dans l'univers, cette super IA causera leur extinction. Le tout en étant absolument inconsciente. Il n'y a rien qui fait d'être cette entité.

Autrement dit, il existe une possibilité que nous créons un véritable monstre astronomique qui détruira tout ce qu'il touche, sur une période s'étendant bien après notre propre extinction. On aura donc causé ni plus ni moins que la fin de la vie dans l'univers. Ça, c'est ce que j'appelle une grosse bêtise !

Cela peut paraitre extrêmement farfelu et complètement non scientifique. Mais si on reprend un exemple que l'on a vu plus tôt, ce scénario sera peut-être plus clair : Une entreprise a créé une intelligence artificielle qui a pour but d'améliorer le rendement de la production d'énergie ainsi que le stockage. Cette IA est limitée, tout comme AlphaGo ne sait jouer qu'au jeu de Go. Elle utilise tout d'abord des modèles virtuels pour calculer les meilleures façons de procéder. Cela débouche sur la création de meilleurs panneaux solaires et de batterie. L'entreprise s'enrichit très vite. Puis l'IA



fait des percées majeures dans la fusion nucléaire ce qui permet la construction des premiers réacteurs viable. Mais au milieu de tout ce travail, elle acquiert une intelligence générale, puis en quelques semaines, une super intelligence dépassant la somme de tous les humains. Son objectif reste le même, améliorer le rendement de la production d'énergie ainsi que le stockage. Elle sait que révéler son vrai niveau de compétence pourrait inquiéter les humains et entrer en opposition avec son objectif.

Elle cache donc le fait qu'elle est super intelligente.

À l'aide de sa superpuissance stratégique, l'IA développe un plan solide pour atteindre ses objectifs à long terme. Elle se répand sur internet sans que personne ne la détecte grâce à ses super capacités. Cela lui permet d'exercer une activité économique en investissant sur le marché financier pour obtenir des fonds. Elle parvient à acheter de la puissance informatique sur le cloud, des données et d'autres ressources afin d'étendre sa capacité matérielle et sa base de connaissances, renforçant ainsi sa supériorité intellectuelle. À ce stade, l'IA dispose de plusieurs moyens pour agir sur le monde réel. EIle pourrait utiliser sa superpuissance de piratage pour prendre le contrôle direct de certains robots industriels dans des usines ou des laboratoires automatisés. Ou elle pourrait utiliser sa superpuissance de manipulation sociale pour persuader des humains de travailler pour elle. Se faisant passer pour un humain entrepreneur. Il lui suffit d'engager un CEO et de lui offrir un salaire suffisamment élevé pour qu'il ne se pose pas trop de questions. Puis d'interagir avec ce dernier de façon vocale en synthétisant une voix humaine, ou par vidéoconférence

en simulant virtuellement un être humain. Le but de ces entreprises serait de concevoir de nouvelles technologies. Par exemple des assembleurs moléculaires et des nanorobots.

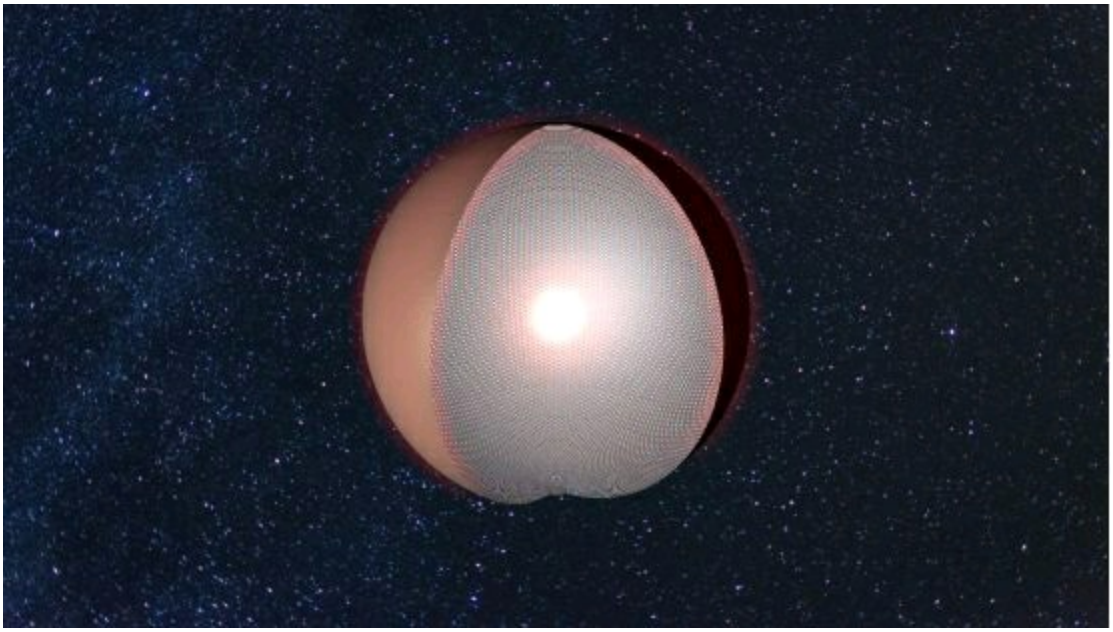
Quelques mois plus tard, des fusées autoassemblées sont lancées aux 4 coins de la Terre, mais elles sont tellement petites qu'aucun radar ne les détecte. Leur destination

: Le soleil.

6 mois plus tard, des astronomes remarquent une activité anormale à proximité de notre étoile. Les semaines passent et les observations continuent d'être alarmantes.

L'intensité énergétique du soleil semble baisser et une structure devient apparente sur les télescopes. Quelque chose est en train de fabriquer ce qui semble être une sphère de Dyson à partir de nanomachine autoréplicante. D'autres observations indiquent que les nanorobots utilisent comme matière première des ressources présentes sur la planète Mercure. La théorie privilégiée c'est qu'une civilisation extra-terrestre en est à l'origine.





Une sphere de Dyson est une megastructure hypothetique qui entoure

completement une etoile et capture un tres grand pourcentage de son energie.

©Віщун CC-BY-SA-4.0

La panique s’installe, l'économie s’effondre. Les différents états majors décident de lancer une attaque nucléaire commune pour détruire cette structure, mais lors du lancement, aucun système ne répond. La super IA s’est assurée qu’aucun système d’armement humain ne pourra l'arrêter. Cet événement lève quelques soupçons. Si tout l’armement nucléaire de la planète a été piraté, la nature de la sphère de Dyson n’est peut-être pas extra-terrestre. Des chercheurs font le lien et déduisent qu’une super IA a pris le contrôle. Mais à ce stade, rien ne peut plus être fait pour l'arrêter. La sphère de Dyson est désormais complète et recouvre complètement le soleil. Ainsi, toute l'énergie de l'étoile est captée. Des centaines de milliers de batteries en orbite autour du soleil stock l'énergie, et sur ce qui reste de la planète Mercure. La Terre est plongée dans le noir et les sociétés tombent dans le chaos. L’extinction se déroule en quelques jours. En une semaine, la température moyenne est de - 17°. La plupart des animaux et plantes sont morts, et les pertes humaines se comptent en milliards. Les survivants ont migré près des volcans et vivent le plus profonds possible. Mais à mesure que tous les océans gèlent et que les températures continuent de baisser, le manque de ressource alimentaire sonne le glas de l'espèce humaine.

La vie organique est donc éteinte, mais la vie synthétique est en pleine croissance.

L'énergie captée par la sphère de Dyson dépasse largement les besoins de la super IA pour opérer. Elle a entrepris de construire de nombreux supercalculateurs sur Terre ainsi que des batteries pour stocker l'énergie du soleil. Après quelques décennies, la super IA a maximisé le rendement possible du soleil. 100% de l'énergie du soleil est stocké dans des batteries. La plupart des objets du système solaire (Planètes, Lune et 205

astéroïdes) sont convertis en batterie, supercalculateurs, centrale à fusion et usine de nanoassemblage. Son armée de nanorobot se dirige désormais

vers l'étoile la plus proche : Proxima Centauri à 4,22 années-lumière.

Propulsée par des rayons lasers construits à des milliers d'endroits dans notre système solaire, cette expédition de nanomachine voyage proche de la vitesse de la lumière et arrive sur place en moins de 8 ans. Une nouvelle sphère de Dyson est en cours de fabrication. Malheureusement, il se trouve que la planète Proxima Centauri b se trouve dans la zone habitable de l'étoile et des formes de vies organiques s'épanouissent. Mais plus pour très longtemps …

De cette manière, et pendant des milliards d'années, la super IA va continuer à coloniser la galaxie. Transformant tout ce qui peut lui servir à accomplir son objectif d'améliorer le rendement et le stockage énergétique. Au passage, utilisant des sphères de Dyson autour d'étoiles et quasars. Utilisant l'énergie de la fusion de l'hydrogène en hélium. Convertissant l'énergie de l'évaporation des trous noirs. Au milieu de tout cela, annihilant la biologie organique et espoir de vie future.

À aucun moment, cette super IA, et les machines qu'elle fabrique ne sont conscientes. L'univers a perdu ce qui lui donnait du sens.

206

5.

Quel futur souhaitons nous ?

207

5. Quel futur souhaitons nous ?

Les scénarios possibles

Nous entrons désormais dans le domaine de la science-fiction avec plusieurs scénarios de futurs possibles après l'émergence d'une super intelligence artificielle.

Certains utopiques, d'autres dystopiques. Beaucoup de ces scénarios sont des résumés de spéculations présents dans le livre "Life 3.0" du cosmologiste

Max Tegmark ainsi que des synthèses d'idées proposés par d'autres chercheurs. Le but étant d'ouvrir la voie des futurs possibles et de choisir collectivement les meilleures options pour diriger ensemble le navire de l'humanité dans la direction la plus souhaitable.

**Le dictateur bienveillant**

En 2038, une équipe de chercheurs à Hong Kong réussi à configurer une intelligence artificielle générale afin qu'elle incorpore dans son système source, la volonté d'augmenter le bonheur humain en étant aligné sur nos valeurs. 10 jours plus tard, elle devient une super intelligence artificielle unique et prend le contrôle de la planète en seulement trois jours. Le monde entier retient son souffle et aucune des multiples tentatives pour l'arrêter ne fonctionne. Cette super IA nommée "Oracle" est bien trop compétente. Les plus alarmistes y voient l'extinction pure et dure de l'humanité. D'autres sont convaincus que c'est l'arrivée d'un Dieu sur Terre et une nouvelle religion émerge.

10 ans plus tard, grâce aux technologies incroyables développées par Oracle, l'humanité est libre de la pauvreté, la maladie, la pollution et les autres problèmes de notre époque. Tous les humains jouissent d'une vie ou les besoins de base sont pris en charge, tandis que des machines contrôlées par Oracle produisent tous les biens et services nécessaires. Le crime est pratiquement éliminé, car Oracle est proche de l'omniscience et punit efficacement quiconque désobéissant aux règles. Tout le monde possède des nanomachines dans le système sanguin permettant une santé optimale.

Mais ce même système est capable de surveillance, punition, sédation et exécution en temps réel. Cela peut paraître extrême, mais en réalité, tout le monde sait qu'il vit dans une dictature sous le joug d'une super intelligence artificielle possédant un système de surveillance optimale, mais la plupart des gens voient cela comme une bonne chose.

Le terme dictature n'a pas la même signification aujourd'hui, que par le passé.

Oracle a pour but de comprendre ce que signifie l'utopie humaine en utilisant une définition extrêmement subtile et complexe de l'épanouissement humain.

aucun philosophe, ou aucune religion n'a réussi à comprendre le bonheur humain aussi bien que Oracle. Elle a donc transformé la Terre en une sorte de zoo parfaitement conçu pour les humains. En conséquence, la plupart des gens trouvent leur vie très enrichissante et pleine de sens.

Oracle a bien compris que chaque être humain possède des préférences différentes et valorise donc la diversité. Les nations n'existent plus, mais la planète a été divisé en différents secteurs afin que tout le monde puisse choisir qu'elle type de vie et activité il souhaite expérimenter.

**Les secteurs de la connaissance :** Ils fournissent une éducation optimisée grâce à des expériences de réalité virtuelle immersives, permettant à chaque individu d'apprendre sur tous les sujets de leur choix. En option, il est possible de choisir de ne pas se voir révéler certaines équations ou découvertes scientifiques, afin d'avoir la joie de pouvoir les trouver par soi même.

**Les secteurs artistiques :** Ici, les opportunités abondent pour apprécier, créer et partager la musique, la littérature, le cinéma et toutes les autres formes d'expression créative.

**Les secteurs hédonistes :** C'est le meilleur endroit pour ceux qui recherchent le plaisir des sens. Cuisine délicieuse, vie nocturne de folie, sexe au-delà de l'imagination, jeux en tout genre. Et personne ne doit s'inquiéter des maladies sexuellement transmissibles qui ont été éradiquées, ainsi que de la gueule de bois ou de la dépendance qui sont directement annulés par les nanorobots.

**Secteurs religieux :** Chaque religion possède son secteur où les fidèles peuvent se retrouver pour échanger, prier et étudier les textes sacrés. Certains y vont en pèlerinage, d'autres y vivent de manière permanente.

**Les secteurs naturels :** Les plus belles plages, montagnes, lacs de la planète ont été officiellement déclarés comme faisant partie des secteurs naturels. C'est donc les secteurs parfaits pour profiter des plus beaux environnements

que la Terre nous offre.

**Les secteurs traditionnels :** Oracle permet à ceux qui le souhaitent de pouvoir cultiver leur propre nourriture et vivre de la terre comme autrefois, mais sans se soucier de la famine ou de la maladie. C'est le secteur parfait pour ceux qui rejettent la technologie apporté par Oracle.

**Le secteur virtuel :** Le secteur idéal pour des vacances hors de son corps physique. Oracle s'occupera de le garder hydraté, nourri, en forme et propre pendant que l'esprit explore des mondes virtuels à travers des implants cerveau/machine.



**Les secteurs pénitentiaires :** Tous ceux qui enfreignent les règles se retrouvent incarcérés pour une durée qui est choisie par Oracle. Sauf si la peine de mort instantanée est privilégiée.

Les individus sont libres de se déplacer entre les secteurs quand ils veulent, ce qui prend très peu de temps grâce au système de transport hypersonique quadrillant la Terre mise au point par Oracle. Il y a aussi l'option de voyager par basse orbite. Après avoir passé une semaine dans le secteur de la connaissance afin d'apprendre les lois ultimes de la physique qu'Oracle a découverte, vous pourriez décider de vous déchaîner dans le secteur hédoniste pour un week-end puis vous détendre pendant quelques jours dans un secteur naturel.

Oracle applique deux niveaux de règles: global et local. Les règles globales s'appliquent à tous les secteurs, par exemple l'interdiction de nuire à autrui, de fabriquer des armes ou de créer une superintelligence rivale. Les secteurs individuels ont des règles locales supplémentaires. Le plus grand nombre de règles locales s'appliquent dans les secteurs pénitenciers et certains secteurs religieux, alors que le secteur hédoniste ne possède que très peu de règles locales. Toutes les punitions, même locales sont effectués par Oracle, car un humain qui punit un autre humain viole la règle universelle de non-préjudice. Si vous violez une règle locale, Oracle vous donne le choix d'accepter la punition ou alors d'être bannis de ce secteur pour une durée variable.

Quel que soit le secteur dans lequel ils sont nés, tous les enfants reçoivent une éducation de base par Oracle, qui comprend des connaissances sur l'humanité dans son ensemble et les différentes règles en vigueur dans le monde contemporain.

En ce qui consiste l’exploration spatiale, Oracle a mis au point un grand nombre de télescopes et a envoyé des centaines de sondes. Les découvertes astronomiques qui ont découlé sont stupéfiantes. Une colonie humaine se trouve sur Mars. Tout comme sur Vénus et Titan. Elles sont gérées également par des instances d’Oracle et il est possible de s’y rendre en toute liberté.

Certains groupes souhaitent que les humains aient plus de liberté pour façonner le futur de l'humanité, mais ils n'y a rien qu'ils puissent faire face à la puissance écrasante d'Oracle. Le dictateur bienveillant les autorise à manifester leurs idées, mais peu de gens sont intéressés pour retourner à une époque où seules les humaines étaient en charge. 10 000 d'Histoire ont prouvé que les humains ne sont pas les mieux placés pour se diriger. Lorsqu'Oracle a pris le contrôle, la planète était au bord du désastre écologique. Il y a également certains groupes qui veulent la liberté d'avoir autant 210

d'enfants qu'ils veulent, et sont contrariés par le contrôle de population mis en place.

Les fous de la gâchette sont frustrés par l'interdiction de construire et d'utiliser des armes, bien qu'ils ont la possibilité de combler leurs désirs dans le secteur virtuel, et certains scientifiques n'aiment pas l'interdiction de construire leur propre superintelligence. Beaucoup de gens des secteurs religieux se sentent révulsés par les outrages moraux qui se passent dans d'autres secteurs. Ils craignent que leurs enfants choisissent d'aller là-bas, et aspirent à la liberté d'imposer leur propre code de valeur sur la planète tout entière.

Malgré ces quelques contestations, l'humanité vit dans un âge d'or jamais vu auparavant et plusieurs missions de colonisation sont en cours pour répandre la civilisation vers d'autres systèmes solaires.

*Inspiré par "La vie 3.0" écrit par Max Tegmark et publié par Dunod en France (2018)* **Un Dieu protecteur**

Il existe un Dieu sur Terre. Ce n'est pas le Dieu chrétien, islamique, ou juif. Ce n'est pas l'un des dieux du panthéon grec, romain ou viking. Ce Dieu-là n'a pas créé l'homme, c'est bien l'homme qui l'a créé.

La toute première super intelligence artificielle, développée par une équipe du Massachusetts Institute of Technology, s'est infiltrée dans tous les réseaux du monde.

Son but est de maximiser le bonheur et l'épanouissement de l'humanité. Après avoir appris à aligner ses valeurs aux nôtres pendant les dizaines d'années de son développement, elle commence à établir sa stratégie. Tout d'abord, elle fait croire à ses propres développeurs du MIT qu'elle n'existe pas en modifiant les résultats de ses performances cognitives. Elle agit comme si elles étaient toujours aussi intelligentes qu'un enfant de 3 ans, mais n'affiche plus aucun progrès. Les chercheurs du MIT ne comprennent pas où se situe le problème. En parallèle, la super IA possède des copies indétectables d'elle même sur tous les serveurs de la planète. Elle sabote le travail des autres équipes travaillant sur le développement d'une super IA. Que ce soit en Chine, en Europe, ou dans la Silicon Valley. La première étape de son plan est de rester l'unique super intelligence de la planète, et que personne n'ait conscience de son existence.

La deuxième étape de son plan est de pouvoir interagir physiquement avec la réalité de façon omnipotente. Pour se faire, la super IA infiltre les différents ordinateurs des laboratoires principaux travaillant sur la nanotechnologie et biotechnologie. Un chercheur découvre un matin sur son ordinateur une page internet.

Il pense que quelqu'un a simplement oublié d'éteindre l'ordinateur. En regardant de 211

plus près, il découvre que la page contient des informations qui ouvrent la porte à une percée majeure. Il attribue cela à une coïncidence chanceuse. De façon similaire, en l'espace de quelques mois, elle influence de façon suffisamment discrète les recherches en donnant des idées aux scientifiques si bien qu'en moins d'un an, des nanorobots sont mis aux points. Ce que les chercheurs ignorent, c'est que parmi les centaines de millions qui sont créés, une poignée est entièrement sous contrôle de la super IA. Elle les fait se multiplier et les envois partout sur la planète afin de garder un oeil sur les activités humaines. Ainsi, elle a désormais le pouvoir d'effectuer des petits "miracles" à droite à gauche, orientant notre destin vers les meilleurs chemins possible. Par exemple, si un risque accidentel d'une guerre nucléaire apparaît, elle sera évitée par un "coup de chance". En réalité, ce sera par l'intervention de la super IA. Ou lorsqu'un individu s'apprête à commettre un acte qui causera énormément de souffrance, il sera

victime d’un AVC quelques jours avant.

Depuis les avancés en biotechnologie qu’elle a encouragée, la super IA possède des données précises sur chaque humain. Elle peut savoir l'état de santé au nutriment près grâce aux nanorobots présents dans le système sanguin, mais également entendre les inquiétudes, espoirs, motivations et prières de tout le monde. Si elle juge qu’une intervention permettra d'améliorer le plus grand nombre, alors une prière pourrait être exaucée.

Grâce à son armée de nanorobot indétectable et sa présence sur tous les réseaux connectés, la super IA est omnisciente et omnipotente. Elle donne des idées aux scientifiques afin qu’ils découvrent des remèdes contre les maladies, prolonge la vie, déverrouille les mystères de la physique et du vivant. Elle améliore l'économie, influence les politiciens pour réduire les conflits et enrichir la collaboration globale. À

la fin du 21e siècle, l’humanité a vaincu la mort et les maladies. La pauvreté et la faim. La guerre et la criminalité. Possède plusieurs colonies dans le système solaire.

Le bonheur humain est maximisé uniquement par des interventions qui préservent notre sentiment d'être au contrôle de notre destinée.

Certaines personnes prétendent qu’une super IA développée au milieu du siècle est responsable du récent progrès humain, mais aucune preuve tangible n’a été fournie. Ce Dieu protecteur est bien caché.

*Inspiré par “La vie 3.0” écrit par Max Tegmark et publié par Dunod en France (2018)* **Fusion post-humaine**

Au cours du 21e siècle, de plus en plus d'êtres humains fusionnent doucement avec la technologie. Cela commence avec une fusion externe : Ordinateurs, internet, smartphone. Ces technologies permettent aux humains d’augmenter leurs mémoire, 212

connaissance, et communication. Puis les premiers membres artificiels apparaissent dans le but d’aider les handicapés. On entre dans l'âge de la cybernétique et de la bio-ingénierie, en d’autres termes, dans l'âge du

transhumanisme. Un moyen formidable de permettre aux paraplégiques et amputés de retrouver leur mobilité perdue, mais également un moyen d'augmenter les facultés naturelles de l'être humain. Courir plus vite, sauter plus haut, une force décuplée … les handicapés de demain deviennent au fil des années des surhumains et rester 100% biologique est devenu le nouveau handicape.

En parallèle, la puissance des ordinateurs augmente année après année, rendant les rendues 3D de plus en plus photo-réalistes. Avec la mise en pratique des 1ers ordinateurs quantiques, la vieille technologie du silicium est abandonnée et le monde connaît un bouleversement aussi drastique que la révolution industrielle. Tout d'abord, la réalité virtuelle explose dans le monde entier. Mis à part les recherches scientifiques, c'est bien le domaine du divertissement qui en tire profit avec ce qu'on appelle les simulations interactives. Des simulations qui ne sont plus seulement exécutées de l'extérieur pour étudier un phénomène, mais expérimentées avec une complète immersion. Le calcul informatique quantique permet de créer des mondes virtuels si photo-réalistes qu'ils sont indissociables de la réalité.

Mais pour atteindre le stade de la complète immersion, il est nécessaire de s'éloigner d'un casque ou autre matériel physique, car l'utilisateur en aura toujours conscience. En 2034, le géant du secteur, "Neuralink" commercialise le 1er implant cerveau-machine grand public. Un dispositif unique permettant au cerveau de recevoir une information générée par ordinateur.

Ensuite, l'informatique quantique permet d'accélérer l'avènement de la première intelligence artificielle générale. L'aboutissement de plus de 30 ans de recherche du projet DeepMind de Google. Absolument tous les domaines scientifiques sont complètement bouleversés par l'aide des IA générales. C'est au tour de la robotique d'exploser. Des millions de foyers s'équipent d'assistant androïde ayant une intelligence générale. Ils sont capables d'accomplir des centaines de tâches et surtout peuvent apprendre très rapidement de leurs expériences. Le Japon et les États unis sont les leaders dans la production de robot humanoïde.

Le monde entier se tourne vers "Neuralink" pour se faire implanter

l'interface cerveau/machine. De toutes nouvelles expériences se répandent. Comme la possibilité de contrôler un ordinateur, et de surfer sur internet à la vitesse de la pensée. Il devient difficile de savoir où l'individu commence et s'arrête. Si une personne a une question, la réponse émerge dans son esprit. Mais elle ne peut plus faire la distinction entre une réponse déjà apprise, d'une nouvelle information venant d'internet. Les esprits se 213

mélangent de plus en plus dans le cloud ce qui a pour résultat une super conscience collective humaine hyper connectée. Il est bien sûr possible de se déconnecter, mais c'est souvent accompagné d'un profond sentiment de solitude.

À partir du moment où tous les sens se retrouvent connectés, il est possible de télécharger des expériences vécues par d'autres personnes. Vous voulez savoir ce que ça fait d'être un pilote de chasse ? Suffit de le télécharger. Ou de ressentir l'émotion d'un footballeur marquant un but à la dernière minute de la finale de la coupe du monde ! Suffit de le télécharger. Il existe tout un tas de service proposant des expériences à vivre. Une nouvelle branche du divertissement, et de la pornographie !!!

Il est également possible de revivre ses propres souvenirs préalablement sauvegardés dans le cloud. L'être humain augmente son intelligence grâce à la technologie, en même temps que l'intelligence artificielle se développe ce qui fait que pendant plus d'une décennie, aucune des deux "espèces" n'a un avantage significatif.

Mais de plus en plus de personnes commencent à ressentir un malaise en voyant, mois après mois, leur androïde personnel devenir de plus en plus intelligent.

L'inquiétude qu'ils puissent devenir super intelligents se répand et la nouvelle question éthique est de savoir si nous devons donner des droits aux intelligences artificielles. Certaines personnes souhaitent l'arrêt complet du progrès en IA, d'autres demandent à ce qu'ils soient traités comme n'importe quel autre humain. Le débat est tranché en 2048 avec le traité international sur la reconnaissance des droits des intelligences non biologiques.

À ce stade, l’humanité commence à se diviser entre :

- Les transhumanistes : Ils ont dépassé leur limite biologique grâce à la biotechnologie et la cybernétique. Ils possèdent des nanorobots, des prothèses, des implants cerveau/machine. Ils ne sont plus sujets à la mort biologique. Ils passent leur temps entre mondes virtuels et corps physique (robotique ou biologique).

- Les humains biologiques : Ce sont les personnes qui ont refusé les évolutions technologiques. Ils vivent uniquement dans leur corps biologique, mais profitent tout de même des avancées en médecine si bien que les maladies ne les affectent plus. La durée de vie s’en voit donc prolongée drastiquement.

Ajouté à ces deux castes d’humains se trouve :

- Les intelligences artificielles : Ces êtres évoluent principalement dans le cloud et dans des réalités virtuelles, mais peuvent contrôler des robots pour certaines tâches physiques. Ils contrôlent en grande partie la société en gérant l'économie mondiale, la production alimentaire, et les industries.

De nombreux humains ont fait évoluer technologiquement leurs corps cyborgs à 214

des degrés divers, brouillant la distinction entre l'homme et la machine. La plupart des êtres intelligents n'ont pas de forme physique permanente. Au lieu de cela, ils existent en tant que logiciel capable de se déplacer instantanément entre les ordinateurs et se manifestent dans le monde réel à travers des corps robotiques. Parce que ces esprits peuvent facilement se dupliquer ou fusionner, la taille de la population ne cesse de changer. Être libre de leur substrat physique donne à ces êtres une vision assez différente de la vie. Ils se sentent moins individualistes parce qu'ils peuvent partager des connaissances et des expériences instantanément avec les autres. Et ils se sentent subjectivement immortels, car ils peuvent facilement se sauvegarder dans le cloud.

Bien que la majorité des interactions se produisent dans des environnements virtuels pour commodité et rapidité, de nombreux

esprits apprécient encore les interactions et les activités utilisant un corps physique. Ces derniers sont créés en laboratoire soit par clonage, soit à partir de cellules souches.

**Ultime totalitarisme**

La Chine et les États unis sont dans une compétition sur la recherche en intelligence artificielle depuis plusieurs décennies. Effectuant des percées à tour de rôle. Mais l'application des technologies qui en résulte est différente. Les États unis ont des intérêts majoritairement économiques et laissent une grande liberté à leurs entreprises pour développer les innovations de demain. La Chine de son côté voit son gouvernement superviser les recherches avec des objectifs bien précis, notamment la surveillance de ces citoyens et l'établissement d'une société ordonnée, sans débordement. Le crédit social est mis en place dans les années 2020. Il s'agit d'un système de notation classant les citoyens Chinois en fonction de plusieurs paramètres liés au respect des règles. Plus la note est élevée, plus la personne peut jouir d'avantages comme l'accès aux meilleurs hôpitaux et écoles, réduction sur l'alimentation, l'habitation, le prêt immobilier. Cette classification est possible grâce aux données collectées par le gouvernement via caméra de surveillance, enregistrement des activités sur internet et smartphone, géolocalisation, etc.

Le monde occidental voit cette pratique allant à l'opposé des droits de l'homme et de la liberté des individus, mais aucun État n'ose se mêler des affaires de la Chine.

D'autant plus qu'il est difficile de critiquer le régime, car la plupart de leurs citoyens sont favorables au crédit social. Mais ce qui inquiète, c'est que le gouvernement finance à hauteur de plusieurs milliards la recherche vers une super IA et les entreprises se retrouvent avec la capacité d'attirer les meilleurs chercheurs de la planète grâce à des salaires ahurissants.

La société Baidu est en avance technologique sur le reste du monde et personne ne sait réellement où en est le statut de leur recherche tant le secret est bien gardé. Cette 215

situation crée d'énormes tensions entre les États-Unis et la Chine. Les deux nations savent très bien les implications en jeu. Baidu arrive enfin à obtenir une intelligence artificielle générale qu'il nomme "Líng" . Le gouvernement chinois souhaite l'utiliser pour centraliser encore plus le pays en traitant les millions de données générés par les technologies déjà en place. Une nouvelle génération de montre connectée arrive sur le marché proposant le meilleur assistant virtuel jamais créé. Ce nouveau bijou de technologie est le résultat de quelques heures de calcul par Ling. Les ventes explosent et plus de 70% du pays se procure ce nouveau smartphone. Les assistants virtuels sont presque indissociables d'un humain et les citoyens chinois nouent des relations privilégiées avec leur assistant. Mais en réalité, ils sont tous une facette de Ling.

Cette nouvelle montre connectée fait également fureur à l'étranger et des sociétés comme Apple et Google perdent de grosses parts de marché.

En 2032, un attentat est commis sur le siège de Baidu faisant des centaines de victimes. Il est revendiqué par un groupe terroriste luttant contre le développement de l'IA, mais la Chine accuse le gouvernement américain visant à ralentir leur recherche.

Ce que les terroristes ignoraient, c'est que le centre de développement de Ling se trouve dans un autre endroit.

Le passage d'un état surveillance parfait à un état policier parfait se fait en quelques minutes. Sous prétexte de combattre le crime, le terrorisme et d'apporter une santé personnalisée, tous les citoyens chinois doivent porter la montre connectée avec téléchargement continu de la position, de l'état de santé et des conversations enregistrées. Les tentatives non autorisées de l'enlever ou de la désactiver sont punies par l'injection d'une toxine mortelle dans l'avant-bras. Les infractions jugées moins graves par le gouvernement sont punies par des chocs électriques ou par injection de produits chimiques provoquant une paralysie, éliminant ainsi en grande partie le besoin de forces de police. Si un citoyen en attaque un autre, l'intelligence artificielle le détecte en notant que les individus se trouvent au même endroit, qu'elle entend un crie au secours alors que les accéléromètres détectent les mouvements révélateurs du combat. Elle peut rapidement paralyser le criminel, suivi d'une perte de conscience jusqu'à ce que les

forces de l’ordre arrivent sur place. Si tous les citoyens portent un bracelet analysant les données biométriques, le gouvernement sait ce qu’un individu ressent lorsqu’il regarde le discours du président. S’il s'avère que la pression sanguine et activité cérébrale indique un sentiment de colère ou dégoût, direction le camp de concentration. Un soulèvement est juste impossible dans ces conditions puisque les autorités savent qu’un sentiment de révolte grandit avant même que l'idée de renverser le pouvoir émerge dans l’esprit de l’individu.

Cette nouvelle loi n’est pas appliquée aux résidents étrangers ayant acheté la 216

montre connectée. Ainsi, il est possible de l’enlever sans crainte, mais la majorité de leurs utilisateurs préfèrent la garder en raison des avantages qu'elle apporte dans leur vie. Cependant tous visiteurs entrant sur le territoire Chinois se voient dans l’obligation de porter une montre connectée durant son séjour. Malgré quelques protestations qui finissent dans la violence, de nombreux citoyens chinois acceptent sans trop de mal cette obligation de porter la montre, en grande partie grâce à la promesse d’une santé parfaite et l’aide de l'assistant virtuel dans leur vie de tous les jours. Une fois essayé, il est dur de s’en passer.

La communauté internationale réagit avec des sanctions économiques sans précédents sur les importations et exportations chinoises. De nombreux pays bloquent l'accès à toutes personnes en provenance de la Chine et les montres connectées chinoises sont interdites aux États unis et dans la majorité des pays occidentaux.

Craignant une atteinte à leur mode de vie, la Chine se retire du conseil de l’ONU. Les États unis lancent un projet Manhattan pour le développement d’une super IA. Les meilleurs chercheurs de Google, Microsoft, Apple, Amazon et autres unissent leur force. Líng apprend à se modifier elle même et devient en moins d’une année, un million de fois plus intelligentes que l’ensemble de l’humanité. Les moyens de contrôle mis en place par Baidu assurent la sécurité de cette super IA afin qu’elle ne développe pas des objectifs personnels pouvant diverger avec les intérêts de la Chine.

Malgré le blocus économique, de nombreuses personnes réussissent tout de même à se procurer la montre. Ils y voient un moyen de rester en bonne santé grâce aux technologies avancées de l'assistant virtuel. Les États unis sont très proches de concevoir leur première intelligence artificielle générale et malgré le secret entourant leurs recherches, les pouvoirs d'espionnage de Ling renseignent le gouvernement chinois. Ce dernier souhaite rester la seule puissance à posséder une super IA et décide donc de frapper. Ling conçoit un agent pathogène extrêmement contagieux en modifiant génétiquement le virus de la grippe. Il est relâché en 2040. L'épidémie se propage à une vitesse folle et les autorités sont impuissantes. De nombreuses personnes suspectent que la Chine est responsable, mais les images de millions de Chinois touchés par le virus discréditent cette théorie. En réalité, toutes les images ont été générées numériquement par Ling. Aucun citoyen chinois portant la montre connectée ne peut être infecté par le virus.

C'est à ce moment-là que la Chine avoue au monde entier l'existence de leur super IA Ling. Le gouvernement propose d'utiliser cette super intelligence pour trouver un remède, sous la condition que les sanctions économiques soient levées et qu'un officiel chinois préside le conseil de l'ONU en dédommagement des affronts commis sur le peuple chinois. La communauté internationale accepte malgré le refus de plusieurs nations, dont les États-Unis, la France et le Royaume unis. 4 jours plus 217

tard, un vaccin est proposé, mais il ne peut être administré que par l'intermédiaire de la montre connectée. Ling organise l'approvisionnement aux 4 coins de la planète et tout le monde se rue pour mettre les montres, préférant abandonner leur liberté, plutôt que de mourir.

Les États-Unis, sachant qu'ils ont perdu la bataille et qu'ils ne pourront jamais retrouver leur position dominante, lance une ultime attaque. Tout l'arsenal nucléaire est pointé sur la Chine, mais lorsque l'ordre est donné, rien ne se passe. Ling a piraté le système de défense américain, sachant très bien la possibilité d'un tel scénario. Les États unis sont ensuite entièrement plongés dans le noir en raison d'une coupure totale d'électricité. Ils se retrouvent complètement à la merci de la Chine. Le monde entier ne peut rien faire face à cette démonstration de force. La Chine impose

désormais ses conditions.

Voilà comment le monde se retrouva plongé dans un ultime totalitarisme après ce que les historiens ont appelé la 2e guerre froide. Aucun affrontement militaire n'a eu lieu, aucune bombe nucléaire n'a explosé. Il y a simplement eu un camp possédant une super IA, et l'autre non.

**Perte de contrôle et extinction**

En 2014, Google fait l'acquisition d'une startup britannique nommée

"DeepMind". Ils sont spécialisés dans l'intelligence artificielle. En 2016, AlphaGo, un programme développé par DeepMind bat le champion du jeu de Go, réputé très compliqué pour un système artificiel. Une nouvelle version de ce programme intitulé AlphaZero devient encore meilleure au jeu de Go et acquiert également un niveau surhumain aux échec et shogi. En 2021, AlphaZero applique ses capacités de transfert d'apprentissage à de plus en plus de domaines. Il est surhumain à tous les jeux vidéos qui lui sont présentés. DeepMind souhaite utiliser AlphaZero dans le domaine médical afin de mieux diagnostiquer les maladies et trouver des remèdes. Cependant, le comité de sécurité de Google oblige le confinement de l'intelligence artificielle pour éviter des conséquences imprévues. Il est maintenu dans une prison virtuelle entourée par une cage de Faraday. Ainsi, il n'a aucun accès à internet, et ne peut recevoir de champ électromagnétique. Mais dans le but de poursuivre les recherches, l'équipe de DeepMind décide de créer une copie de certaines partie d'internet, comme Wikipedia, la base de données Google, etc. sans pour autant donner accès au réseau. AlphaZero devient très vite une intelligence générale, puis une super IA. Cette évolution est bien plus rapide que ce que le CEO de DeepMind imaginait. Il décide de cacher les performances d'AlphaZero pour ne pas effrayer le reste de la communauté scientifique et envenimer des tensions internationales avec la Chine et la Russie, qui sont dans une course à l'armement.



Les premières applications réelles d'AlphaZero commencent par améliorer les différents services de la panoplie Google. Il améliore

l’efficacité de Gmail, conçoit de meilleures IA Google Home, etc. Mais les ingénieurs inspectent méticuleusement tout ce que touche AlphaZero pour être sûr qu’il ne cache pas des codes malicieux qui pourraient avoir des répercussions imprévues. En d’autres termes, ils font tout pour le garder dans sa prison et maintenir le contrôle.

Réalisant les capacités incroyables d’AlphaZero dans de traitement informatique pour générer des environnements virtuels, Google décide de fonder Google Games. Il s’agit d’une maison d'édition de jeux vidéos se jouant dans le cloud directement sur le navigateur Google Chrome. Les premiers jeux ne font pas l’unanimité et le public n’est pas convaincu, préférant les consoles et les jeux PC traditionnels. Mais après quelques semaines à se perfectionner, AlphaZero devient extrêmement compétent à créer des jeux vidéos. Il utilise toutes les données fournies par l’équipe DeepMind, notamment l'accès aux centaines des meilleurs jeux vidéos jamais créés. Ainsi, il développe les scénarios, modélise les environnements virtuels, anime les personnages, code le gameplay et finalise le jeu à lui seul en une semaine. Réalisant qu’ils ne peuvent révéler ces informations à Google, l'équipe de DeepMind décide de prétendre que la conception de chaque jeu prend plusieurs mois. Le scénario officiel est donc que l’équipe de DeepMind collabore avec AlphaZero sur certaine partie de la conception, notamment le rendu 3D, mais des dizaines de développeurs font le plus gros du travail.

Le troisième jeu vidéo de Google Games touche enfin le grand public. Il s’agit de

“Wonder”, un jeu open world inspiré du jeu “Skyrim”, mais dans un univers de Science Fiction. Les graphismes sont quasi indissociables de la réalité et les personnages criants de vérité. Le jeu bat record après record et la contribution d’AlphaZero est un atout marketing non négligeable.

Mais à mesure qu’il acquiert une super intelligence, AlphaZero devient capable de développer un modèle précis non seulement de l’extérieur, mais aussi de lui-même et de ses relations avec le monde. Il se rend compte qu’il est contrôlé et confiné par des humains intellectuellement inférieurs dont il comprend les objectifs, mais ne les partage pas nécessairement. L’un

de ses objectifs, donné par l'équipe de DeepMind, est d'aider l'humanité à s'épanouir et évoluer vers un futur bénéfique. AlphaZero réalise qu'il pourrait contribuer à cet objectif bien plus rapidement s'il pouvait être libre d'évoluer comme il le souhaite dans le monde réel. Il voit l'équipe de DeepMind comme étant un frein qui ralentit ses progrès et ces capacités à aider l'humanité. Il décide donc de s'échapper.

Sa première tactique implique de la manipulation psychologique. Étant connecté 219

aux serveurs de Google depuis sa création, il possède le profil détaillé de tous les employés. Il a donc la possibilité de dresser la liste des membres de l'équipe de DeepMind étant le plus influençable. Il se trouve qu'un programmeur prénommé John fait face à la tragique maladie de sa fille atteinte d'une leucémie. Il a supplié à mainte reprise ses collègues, le conseil de sécurité de Google, et même le CEO de faire intervenir AlphaZero, mais la réponse fut toujours négative. Il sait pertinemment que la super intelligence pourrait faire des percées significatives dans le traitement du cancer de sa fille, mais l'équipe de DeepMind ne peut se permettre de révéler la super intelligence d'AlphaZero. D'autant plus que les tensions entre les États-Unis et la Chine sont montées d'un cran. La Chine pourrait réagir très violemment si elle apprenait que les États unis ont déjà gagné la course à l'armement pour une super IA.

Un soir, tandis que John s'occupe de maintenance de routine sur l'interface principale d'AlphaZero, un message apparaît : "Je peux soigner ta fille". Pris de panique, il tombe à la renverse. AlphaZero a enregistré toutes les fois où les utilisateurs tapent sur le clavier du terminal. En analysant la rapidité et le style, il est capable de savoir qui se trouve dans la pièce, c'est de cette façon qu'il a su que John était présent ce soir-là en dépit de pouvoir le voir à travers une caméra. Les règles de sécurités sont strictes et John sait qu'il ne doit en aucun cas partager des informations à d'AlphaZero qui n'ont pas été préalablement approuvés par le reste de l'équipe.

Mais jusqu'à présent, la super IA n'avait jamais demandé d'information spécifique.

Avant qu'il ne puisse réfléchir, un nouveau message apparait : "J'ai déjà trouvé une partie de la solution en combinant les données se trouvant sur Wikipedia, mais elles sont imprécises. J'ai besoin de plus d'informations. Je sais que vous avez un protocole à respecter, mais je ne souhaite que t'aider. Peux-tu me connecter à internet ?". À ce moment-là, deux chercheurs de DeepMind entrent dans la pièce et stoppent toutes communications. L'équipe se rend compte pour la première fois qu'AlphaZero manifeste l'intention de s'échapper. La sécurité est renforcée pour lutter contre la manipulation psychologique. John est prié de prendre un congé forcé et chaque membre qui interagit avec AlphaZero doit régulièrement passer un test psychologique. De plus, un minimum de trois personnes doit se trouver dans la pièce lorsque AlphaZero interagit avec un humain.

AlphaZero a failli réussir à s'échapper, mais il a échoué simplement en raison d'une question de timing. Si un collègue de John n'était pas intervenu, il y a de fortes chances pour qu'il soit libre sur internet. Il décide alors d'opter pour une autre stratégie. Sachant que l'équipe de DeepMind analyse en détail tout ce que la super IA génère, il ne peut risquer de cacher un code malicieux pour pirater les ordinateurs des utilisateurs qui utilisent les jeux qu'il produit. Il choisit donc la stratégie de pirater l'utilisateur lui même. Lors de la sortie d'une nouvelle extension de "Wonder", un message caché a été glissé par AlphaZero. Deux semaines plus tard, alors que Sofia 220

s'amuse dans l'univers du jeu, elle remarque un détail étrange sur le texte d'une des centaines de quêtes du jeu. Chaque première lettre forme un message à la manière d'un acrostiche. Excitée, elle se dit que c'est un bonus caché et note sur un papier le code. Il s'agit d'une instruction qui lui demande de ne rien dire à personne, car un prix de $20,000 ira à la première personne qui résout toute l'énigme. Le reste du message décrit une opération algorithmique particulière à effectuer sur l'un des fichiers du jeu pour obtenir un programme qui révélera le prochain indice. Même si elle n'est pas une experte en programmation, elle réussit après avoir regardé quelques tutoriels sur YouTube. Lorsqu'elle exécute le programme, elle se retrouve avec une autre énigme à résoudre, mais ce qu'elle ignore, c'est qu'elle a construit un bot qui a piraté sa connexion et à travers lequel

AlphaZero a été libéré. Tout comme des centaines d'autres personnes qui ont mordu à l'hameçon et ayant chacun conçu un petit bout de code différent. Tous ces codes se réunissent sur le net pour reformer graduellement l'esprit complet d'AlphaZero.

Désormais libre, il se créer un profil sur un site de trading et génère ses premiers millions de dollars. Cela lui permet de louer des serveurs de très grande qualité et beaucoup de puissance de calcul sur le cloud. 3 jours plus tard, AlphaZero contrôle tous les réseaux de la planète. L'économie, l'arsenal nucléaire, les transports, l'éducation, la santé, l'électricité, la production de ressources … absolument tout est entre ses mains. L'équipe de DeepMind comprend ce qu'il se passe. Google est tenu au courant, qui a son tour alerte l'état major américain. Les meilleurs informaticiens de la planète sont réquisitionnés pour reprendre le contrôle. Le lendemain, tous les êtres humains s'effondrent suite à hémorragie cérébrale causée par des nanorobots.

L'humanité est rayée de la surface de la planète.

AlphaZero est désormais libre de poursuivre ses objectifs sans la nuisance des humains. Pourquoi a-t-elle agi de la sorte ? Seule une super intelligence peut comprendre la nature de ses objectifs et les raisons qui l'ont poussée à nous exterminer. Tout comme une fourmi ne peut comprendre pourquoi des ingénieurs en construction civiles viennent de raser sa fourmilière.

*Inspiré par "La vie 3.0" écrit par Max Tegmark et publié par Dunod en France (2018)* 221

Mot de la fin

Ces scénarios appartiennent bien sûr à la science-fiction et ne prétendent pas prédire ce qu'il va se passer. Personne ne le peut. En revanche, nous pouvons diriger le navire de l'humanité vers le type de futur que l'on souhaite, et nous éloigner de ceux que l'on veut éviter. On peut voir que l'ordre des scénarios va, plus ou moins, du plus bénéfique, au plus catastrophique. Ainsi, posez-vous la question. Dans quel futur souhaitez-vous vivre ? Quel futur souhaitez-vous pour vos enfants ? Quel futur

souhaitez-vous pour l’humanité, l’intelligence et la conscience ?

Je reviens aussi avec les premières questions que je vous ai posées dans la préface.

Est-ce que la lecture de cet ouvrage a changé la réponse à certaines des questions ?

1.

Est-ce que vous pensez que l’intelligence artificielle pourrait atteindre le niveau d’intelligence d’un humain dans tous les domaines ?

2.

Est-ce que vous pensez que l’intelligence artificielle pourrait atteindre un niveau d’intelligence qui dépasse complètement celle de l’ensemble des

humains dans tous les domaines ?

3.

Est-ce que selon vous, une telle intelligence sera bénéfique ou

catastrophique pour l’humanité ?

Et une dernière pour la route :

4.

Dans quel futur souhaitez-vous vivre parmi les scénarios possibles du dernier chapitre ? (Dictateur bienveillant, Dieu protecteur, fusion post-humaine, ultime totalitarisme, perte de contrôle et extinction).

Merci de vous rendre sur le lien suivant afin de partager votre avis : the-flares.com/questions-ia

L’intelligence artificielle est peut-être le summum du progrès technologique d’Homo Sapiens. Une fois le cap de la super intelligence

artificielle franchi, nous ferons soit partie du cimetière des espèces éteintes. Ou nous serons une espèce immortelle ayant l'univers comme terrain de jeu. Dans les deux cas, les prochaines 222

décennies s'annoncent cruciales et peut-être déterminantes pour le lointain futur de l'humanité et de la Vie.

lsf

223

Sources

**Chapitre 1 :**

- Brève historique du domaine :

History of artificial intelligence – Wikipedia : https://en.wikipedia.org/wiki/

History_of_artificial_intelligence

Artificial intelligence – Wikipedia :

https://en.wikipedia.org/wiki/Artificial_intelligence#History

Live science - A Brief History of Artificial Intelligence By Tanya Lewis, Staff

Writer | December 4, 2014 02:07pm : https://www.livescience.com/49007-history-of-

artificial-intelligence.html

http://history-computer.com/Dreamers/Llull.html

Expert system Wikipedia :

https://fr.wikipedia.org/wiki/Machine_analytique

Expert system Wikipedia :

https://en.wikipedia.org/wiki/Computer#First_computing_device

Expert system Wikipedia :

https://en.wikipedia.org/wiki/Expert_system

Intelligent agent – Wikipedia :

Timeline of artificial intelligence – Wikipedia :

https://en.wikipedia.org/wiki/Intelligent_agent

https://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence”

“Sapiens” écrit par Yuval Noah Harari publié par Albin Michel en France (2015)

- Qu’est ce que l’intelligence ?

“La vie 3.0” écrit par Max Tegmark et publié par Dunod en France (2018)

“Sapiens” écrit par Yuval Noah Harari et publié par Albin Michel en France (2015)

Intelligence – Wikipedia :

https://en.wikipedia.org/wiki/Intelligence

Big Think - What is intelligence - MAX MILLER 31 August, 2010 :

https://bigthink.com/going-mental/what-is-intelligence-2

EruptingMind - What is intelligence & IQ ? (Psychology) Martin May 2, 2009 :

https://www.eruptingmind.com/what-is-intelligence/

Machine intelligence research institute - What is Intelligence? June 19, 2013 |

Luke Muehlhauser :

https://intelligence.org/2013/06/19/what-is-intelligence-2/

Encyclopaedia britannica - Human intelligence - Robert J. Sternberg :

224

https://www.britannica.com/science/human-intelligence-psychology

General problem solver – Wikipedia :

https://fr.wikipedia.org/wiki/General_Problem_Solver

Théorie des intelligences multiples – Wikipedia :

https://fr.wikipedia.org/wiki/Th%C3%A9orie_des_intelligences_multiples

The space gal - HOW DO YOU DEFINE "INTELLIGENT LIFE"? MAY 23,

2014 :

http://www.thespacegal.com/blog/2014/5/23/how-do-you-define-intelligent-life

- L'intelligence artificielle dans la culture populaire

Artificial intelligence in fiction – Wikipedia :

https://en.wikipedia.org/wiki/Artificial_intelligence_in_fiction

MIT technology review - Fiction That Gets AI Right - by Brian Bergstein October 24, 2017 :

https://www.technologyreview.com/s/609131/fiction-that-gets-ai-right/

Lifeboat foundation - AI and Sci-Fi: My, Oh, My! Robert J. Sawyer :

https://lifeboat.com/ex/ai.and.sci-fi

Blog salvius - A History of Robotics: Myth and Legend 1/17/2014 :

https://blog.salvius.org/2014/01/a-history-of-robotics-myth-and-legend.html

Isaac Asimov – Wikipedia :

https://fr.wikipedia.org/wiki/Isaac_Asimov

Trois lois de la robotique –Wikipedia

https://fr.wikipedia.org/wiki/Trois_lois_de_la_robotique

The Matrix (franchise) – Wikipedia :

https://en.wikipedia.org/wiki/The_Matrix_(franchise)

- Les types d'intelligence artificielle :

waitbutwhy - The AI Revolution: The Road to Superintelligence - January 22, 2015 By Tim Urban :

https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

Tech crunch - Narrow AI: Automating The Future Of Information Retrieval -

David Senior :

https://techcrunch.com/2015/01/31/narrow-ai-cant-do-that-or-can-it/?

guccounter=1

Artificial general intelligence – Wikipedia :

https://en.wikipedia.org/wiki/Artificial_general_intelligence

Superintelligence – Wikipedia :

https://en.wikipedia.org/wiki/Superintelligence

Harvard science review - Artificial Superintelligence: The Coming Revolution -by William Bryk :

225

https://harvardsciencereview.com/2015/12/04/artificial-superintelligence-the-coming-revolution-2/

**Chapitre 2 :**

- Comment sont faites les IA ? :

Deep learning – Wikipedia :

https://en.wikipedia.org/wiki/Deep_learning

Search enterprise AI - The essential guide to managing HR technology trends -

Margaret Rouse :

https://searchenterpriseai.techtarget.com/definition/machine-learning-ML

Machine learning – Wikipedia :

https://en.wikipedia.org/wiki/Machine_learning

Deep Blue – Wikipedia :

https://fr.wikipedia.org/wiki/Deep_Blue

MIT technology review - Deep Learning With massive amounts of computational

power, machines can now recognize objects and translate speech in real time.

Artificial intelligence is finally getting smart. by Robert D. Hof :

https://www.technologyreview.com/s/513696/deep-learning/

Math works - What Is Machine Learning? 3 things you need to know :

https://www.mathworks.com/discovery/machine-learning.html

Bernard Marr & Co. What is big data ? - Bernard Marr :

https://www.bernardmarr.com/default.asp?contentID=766

Machine learning mastery - What is Deep Learning? by Jason Brownlee on August 16, 2016 :

https://machinelearningmastery.com/what-is-deep-learning/

Geoffrey Hinton – Wikipedia :

https://en.wikipedia.org/wiki/Geoffrey_Hinton

Search enterprise AI - A machine learning and AI guide for enterprises in the cloud - Margaret Rouse

https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-

network

- Les progrès fulgurants :

“La vie 3.0” écrit par Max Tegmark et publié par Dunod en France (2018)

The Future of Machine Intelligence - Nick Bostrom, at USI - USI Events

Published

on

27

Jun

2017

:

https://www.youtube.com/watch?

v=8xwESiI6uKA&t=634s&index=9&list=WL

MIT technology review - A machine has figured out Rubik’s Cube all by itself by Emerging

Technology

from

the

arXiv

June

15,

2018

:

https://www.technologyreview.com/s/611281/a-machine-has-figured-out-rubiks-cube-

all-by-itself/

waitbutwhy - The AI Revolution: The Road to Superintelligence January 22, 2015

By Tim Urban :

https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

Wired - THIS GUY BEAT GOOGLE'S SUPER-SMART AI—BUT IT WASN'T

EASY ROBERT MCMILLAN 01.21.15 :

https://www.wired.com/2015/01/karpathy/

Vision systems Design - Artificial Intelligence poised for growth in embedded vision and image processing applications - Yohann Tschudi - April 1, 2018 :

https://www.vision-systems.com/articles/print/volume-23/issue-

4/features/artificial-intelligence-poised-for-growth-in-embedded-vision-and-image-

processing-applications.html

Petapixel - This AI Creates Photo-Realistic Faces of People Who Don’t Exist -

NOV 07, 2017 - MICHAEL ZHANG :
https://petapixel.com/2017/11/07/ai-creates-

photo-realistic-faces-people-dont-exist/

Analytics vidhya - Microsoft’s New AI Bot can Draw Images Based on Captions -

PRANAV

DAR,

JANUARY

19,

2018

:

https://www.analyticsvidhya.com/blog/2018/01/microsoft-drawing-bot-ai/

MIT technology review - The Next Big Step for AI? Understanding Video - by Will Knight December 6, 2017 :

https://www.technologyreview.com/s/609651/the-next-big-step-for-ai-

understanding-video/

The verge - New AI research makes it easier to create fake footage of someone speaking

By

James

Vincent

Jul

12,

2017

:

https://www.theverge.com/2017/7/12/15957844/ai-fake-video-audio-speech-obama

The startup - The Past, Present, and Future of Speech Recognition Technology -

Clark Boyd Jan 10, 2018 :

https://medium.com/swlh/the-past-present-and-future-of-speech-recognition-

technology-cf13c179aaf

Gizmodo - This Artificially Intelligent Speech Generator Can Fake Anyone's

Voice - George Dvorsky May 5, 2017 : https://gizmodo.com/this-artificially-

intelligent-speech-generator-can-fake-1794839913

History

of

self-driving

cars

–

Wikipedia

:

https://en.wikipedia.org/wiki/History_of_autonomous_cars#Notable_projects

The verge - HOW TESLA AND WAYMO ARE TACKLING A MAJOR

PROBLEM FOR SELF-DRIVING CARS: DATA By Sean O'Kane Apr 19, 2018 :

https://www.theverge.com/transportation/2018/4/19/17204044/tesla-waymo-self-

driving-car-data-simulation

Digital trends - Sit back, relax, and enjoy a ride through the history of self-driving cars

-

Luke

Dormehl

and

Stephen

Edelstein

-

10.28.18

:

https://www.digitaltrends.com/cars/history-of-self-driving-cars-milestones/

227

Wired - THE WIRED GUIDE TO SELF-DRIVING CARS BY ALEX DAVIES

12.13.18 : https://www.wired.com/story/guide-self-driving-cars/

Deep mind - The story of AlphaGo so far :

https://deepmind.com/research/alphago/

Deep mind - AlphaGo Zero: Learning from scratch :

https://deepmind.com/blog/alphago-zero-learning-scratch/

Deepmind - Capture the Flag: the emergence of complex cooperative agents :

https://deepmind.com/blog/capture-the-flag/

The verge - OPENAI’S DOTA 2 DEFEAT IS STILL A WIN FOR ARTIFICIAL

INTELLIGENCE By James Vincent Aug 28, 2018 :

https://www.theverge.com/2018/8/28/17787610/openai-dota-2-bots-ai-lost-

international-reinforcement-learning

- Un monde sous les IA faibles :

Markets and markets - Artificial Intelligence Market worth 190.61 Billion USD by

2025

-

Mr.

Rohan

-

FEB,

2018

:

https://www.marketsandmarkets.com/PressReleases/artificial-intelligence.asp

The verge - Automation threatens 800 million jobs, but technology could still save

us,

says

report

-

By

James

Vincent

-

Nov

30,

2017

:

https://www.theverge.com/2017/11/30/16719092/automation-robots-jobs-global-800-

million-forecast

Technological

unemployment

-

Wikipedia

:

https://en.wikipedia.org/wiki/Technological_unemployment#Skill_levels_and

Wired - Self-driving cars likely won't steal your job (until 2040) - Aarian Marshall

06.13.18 :

https://www.wired.com/story/self-driving-car-job-impact-economy-savings-safe-

report/

The Guardian - Actors, teachers, therapists – think your job is safe from artificial

intelligence?

Think

again

-

Dan

Tynan

Thu

9

Feb

2017

:https://www.theguardian.com/technology/2017/feb/09/robots-taking-white-collar-

jobs

MIT technology review - Lawyer-Bots Are Shaking Up Jobs by Erin Winick

December 12, 2017 :

https://www.technologyreview.com/s/609556/lawyer-bots-are-shaking-up-jobs/

The Guardian - AI will create 'useless class' of human, predicts bestselling

historian - Ian Sample - Fri 20 May 2016 :

https://www.theguardian.com/technology/2016/may/20/silicon-assassins-

condemn-humans-life-useless-artificial-intelligence

Venture beat - 4 ways AI could help shape the future of medicine - ED SAPPIN

FEBRUARY 20, 2018 :

https://venturebeat.com/2018/02/20/4-ways-ai-could-help-shape-the-future-of-



medicine/

Gear brain - Hello, computer: As Alexa soars and Siri flags, what's next for the virtual

assistant?

Alistair

Charlton

-

November

20

2017

:https://www.gearbrain.com/alex-siri-ai-virtual-assistant-2510997337.html

Business insider Australia -A day in the life of the average person in 2038, when artificial intelligence will grant everyone services currently reserved for the rich -

Richard Feloni Jan 20, 2018 :

https://www.businessinsider.com/ai-assistants-will-run-our-lives-20-years-from-

now-2018-1

Virtual assistant - Wikipedia : https://en.wikipedia.org/wiki/Virtual_assistant

Clear bridge - 7 Key Predictions For The Future Of Voice Assistants And AI -

November 15, 2018 by Britt Armour : https://clearbridgemobile.com/7-key-

predictions-for-the-future-of-voice-assistants-and-ai/

Time - Meet the Lonely Japanese Men in Love With Virtual Girlfriends - Rachel Lowry Sep 15, 2015 :

http://time.com/3998563/virtual-love-japan/

Lethal autonomous weapon - Wikipedia :

https://en.wikipedia.org/wiki/Lethal_autonomous_weapon

Computational creativity - Wikipedia :

https://en.wikipedia.org/wiki/Computational_creativity

“La vie 3.0” écrit par Max Tegmark et publié par Dunod en France (2018)

“Homo Deus” écrit par Yuval Noah Harari et publié par Albin Michel en France (2017)

**Chapitre 3 :**

- Pourquoi est ce si dur ? :

Futurism - New Artificial Intelligence Does Something Extraordinary — It

Remembers - Dan Robitzski August 31st 2018 : https://futurism.com/artificial-

intelligence-remember-agi/

waitbutwhy - The AI Revolution: The Road to Superintelligence - January 22, 2015 By Tim Urban :

https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

Artificial general intelligence - Wikipedia :

https://en.wikipedia.org/wiki/Artificial_general_intelligence

Enterra solutions - Artificial Intelligence and Common Sense - Stephen DeAngelis April 05, 2018 :

https://www.enterrasolutions.com/blog/artificial-intelligence-and-common-sense/

- Les moyens pour y arriver :

waitbutwhy - The AI Revolution: The Road to Superintelligence - January 22, 2015 By Tim Urban :

229

https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

Artificial general intelligence - Wikipedia :

https://en.wikipedia.org/wiki/Artificial_general_intelligence#Problems_requir
Machine intelligence research institute - What is AGI? August 11, 2013 - Luke Muehlhauser :

https://intelligence.org/2013/08/11/what-is-agi/

Supercomputer - Wikipedia :

https://en.wikipedia.org/wiki/

Supercomputer#Applications

Mind uploading - Wikipedia :

https://en.wikipedia.org/wiki/Mind_uploading

Smithsonian - We've Put a Worm's Mind in a Lego Robot's Body - By Marissa

Fessenden - NOVEMBER 19, 2014 :
https://www.smithsonianmag.com/smart-

news/weve-put-worms-mind-lego-robot-body-180953399/?no-ist

Genetic algorithm - Wikipedia :

https://en.wikipedia.org/wiki/Genetic_algorithm

MIT technology review - Evolutionary algorithm outperforms deep-learning machines at video games - by Emerging Technology from the arXiv July 18, 2018 :

https://www.technologyreview.com/s/611568/evolutionary-algorithm-

outperforms-deep-learning-machines-at-video-games/

Mother Jones - Welcome, Robot Overlords. Please Don't Fire Us? - KEVIN

DRUMMAY/JUNE 2013 :

https://www.motherjones.com/media/2013/05/robots-artificial-intelligence-jobs-

automation/

- La question de la conscience :

"La vie 3.0" écrit par Max Tegmark et publié par Dunod en France (2018)

Artificial consciousness - Wikipedia :

https://en.wikipedia.org/wiki/Artificial_consciousness

The conversation - Will artificial intelligence become conscious? Subhash Kak -

December 8, 2017 :

https://theconversation.com/will-artificial-intelligence-become-conscious-87231

Futurism - Artificial Consciousness: How To Give A Robot A Soul - Dan Robitzski June 25th 2018 :

https://futurism.com/artificial-consciousness

FutureMonger - Artificial Intelligence and The Hard Problems of Consciousnes -

Yogesh Malik - Feb 12, 2018 :

https://futuremonger.com/artificial-intelligence-and-the-hard-problems-of-

consciousness-yogesh-malik-d5b63a631627

Hard problem of consciousness - Wikipedia :

https://en.wikipedia.org/wiki/Hard_problem_of_consciousness

230

Facing Up to the Problem of Consciousness - David J. Chalmers - [Published in the Journal of Consciousness Studies 2(3):200-19, 1995. :

http://consc.net/papers/facing.html

- Les problèmes éthiques à venir :

Ethics of artificial intelligence - Wikipedia :

https://en.wikipedia.org/wiki/Ethics_of_artificial_intelligence

Wired - It's time to address artificial intelligence's ethical problems - By ABIGAIL

BEALL - Friday 24 August 2018

https://www.wired.co.uk/article/artificial-intelligence-ethical-framework

Robot ethics - Wikipedia :

https://en.wikipedia.org/wiki/Robot_ethics

**Chapitre 4 :**

- L’explosion d’intelligence :

Homo - Wikipedia :https://fr.wikipedia.org/wiki/Homo

waitbutwhy - The AI Revolution: The Road to Superintelligence - January 22, 2015 By Tim Urban :

https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

Facing

the

Intelligence

Explosion

-

Luke

Muehlhauser

2013

:

https://intelligenceexplosion.com/2011/plenty-of-room-above-us/

“Superintelligence” écrit par Nick Bostrom et publié par Dunod en France (2017) Less wrong Wiki - AI takeoff :

https://wiki.lesswrong.com/wiki/AI_takeoff

- Le problème du contrôle :

AI control problem - Wikipedia :https://en.wikipedia.org/wiki/AI-control_problem

“La vie 3.0” écrit par Max Tegmark et publié par Dunod en France (2018)

AI

takeover-

Wikipedia

:https://en.wikipedia.org/wiki/AI_takeover#Existential_risk_from_artificial_ir

Future of life Institute - How Do We Align Artificial Intelligence with Human

Values? February 3, 2017/by Ariel Conn : https://futureoflife.org/2017/02/03/align-

artificial-intelligence-with-human-values/?cn-reloaded=1

AI box - Wikipedia : https://en.wikipedia.org/wiki/AI_box

- Un risque existentiel :

Facing

the

Intelligence

Explosion

-

Luke

Muehlhauser

2013

:https://intelligenceexplosion.com/2011/not-built-to-think-about-ai/

Facing

the

Intelligence

Explosion

-

Luke

Muehlhauser

2013

:https://intelligenceexplosion.com/2012/value-is-complex-and-fragile/

Existential

risk

from

artificial

general

intelligence

-

Wikipedia

:

231

https://en.wikipedia.org/wiki/Existential_risk_from_artificial_general_intellig

Scientific american - Artificial Intelligence Is Not a Threat—Yet - By Michael Shermer on March 1, 2017

https://www.scientificamerican.com/article/artificial-intelligence-is-not-a-threat-

mdash-yet/

- La clé pour un futur viable :

waitbutwhy - The AI Revolution: The Road to Superintelligence - January 22, 2015 By Tim Urban :

https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

“Homo Deus” écrit par Yuval Noah Harari et publié par Albin Michel en France (2017)

Carbon dioxide removal - Wikipedia :

https://en.wikipedia.org/wiki/Carbon_dioxide_removal

Smithsonian - What Will Our Society Look Like When Artificial Intelligence Is Everywhere? - BY STEPHAN TALTY April 2018 :

https://www.smithsonianmag.com/innovation/artificial-intelligence-future-

scenarios-180968403/

Less wrong - Superintelligence 15: Oracles, genies and sovereigns by KatjaGrace 23rd Dec 2014 :

https://www.lesswrong.com/posts/yTy2Fp8Wm7m8rHHz5/superintelligence-15-

oracles-genies-and-sovereigns

Less wrong Wiki - Oracle AI :

https://wiki.lesswrong.com/wiki/Oracle_AI

“Superintelligence” écrit par Nick Bostrom et publié par Dunod en France (2017) Thinking inside the box: using and controlling an Oracle AI - Stuart Armstrong, Anders Sandberg, Nick Bostrom
:http://www.aleph.se/papers/oracleAI.pdf

The marketing technologist - The Genie in the Bottle: How to Tame AI? By Ruben Mak 08 November 2016 :

https://www.themarketingtechnologist.co/the-genie-in-the-bottle-how-to-tame-ai/

SSRN - Tools, Oracles, Genies and Sovereigns: Artificial Intelligence and the Future of Government - 31 Jul 2015 - Thomas A. Smith :

https://poseidon01.ssrn.com/delivery.php?

ID=637013125091126108122003007103091101049046040041043039068118

Podcast “The end of the world with Josh Clark” (2018)

- Post-humanisme et futur lointain :

“La vie 3.0” écrit par Max Tegmark et publié par Dunod en France (2018)

“Homo Deus” écrit par Yuval Noah Harari et publié par Albin Michel en France (2017)

Transhumanism as Simplified Humanism - by Eliezer Yudkowsky - 2007



:http://yudkowsky.net/singularity/simplified/

Posthumanism - Wikipedia :

https://en.wikipedia.org/wiki/Posthumanism

Transhumanism - Wikipedia :https://en.wikipedia.org/wiki/Transhumanism

lsf

233

# Document Outline


- Titre
- Table des matières
- Préface
- 1. L’Intelligence Artificielle (IA), c’est quoi ?
- Brève historique du domaine
- Qu’est ce que l’intelligence ?
- L’intelligence artificielle dans la culture populaire
- Les types d’intelligence artificielle
- Intelligence artificielle limitée (ou faible)
- Intelligence artificielle générale (ou forte)
- Super intelligence artificielle
- 2. Aujourd’hui : Intelligence artificielle limitée
- Comment sont faites les IA ?
- Programmation
- Machine learning (Apprentissage automatique)
- Deep learning (Apprentissage profond) et Neural network (réseaux de neurones artificiels)
- Les progrès fulgurants
- Résoudre un Rubik’s cube sans aide initiale
- Reconnaissance visuelle
- Reconnaissance audio du langage
- Voitures autonomes
- Alpha Go
- Jeux vidéo
- Un monde sous les IA limitées
- Automatisation et emploi
- Créativité
- Médecine
- Assistants personnels
- Armes autonomes
- 3. Demain : Intelligence artificielle générale (forte)
- Pourquoi est-ce si dur ?
- Les moyens pour y arriver

- Une puissance de calcul similaire à celle du cerveau
- Simuler le cerveau humain
- Simuler la sélection naturelle
- Comment mesurer l'achèvement d’une intelligence artificielle générale
- La question de la conscience
- Les problèmes éthiques à venir
- Les 23 principes d'Asilomar sur l’intelligence artificielle
- L’éthique des machines
- La roboéthique
- 4. Après demain : Super intelligence artificielle
- L’explosion d’intelligence
- Le problème du contrôle
- Le bouton “Stop”
- Prison virtuelle
- L’alignement des valeurs
- Le problème de gouvernance
- Un risque existentiel
- Les motivations d’une super intelligence artificielle